Facultad de Ingeniería • UNCUYO ▪ Red de Universidades de Carreras de Informática RedUNCI

WICC 2025

LIBRO DE ACTAS

XXVII

Workshop de Investigadores en Ciencias de la Computación

Mendoza • Abril 2025

UNCUYO UNIVERSIDAD NACIONAL DE CUYO | FACULTAD DE INGENIERÍA | RedUNCI

Workshop de investigadores en ciencias de la computación WICC 2025 : libro de actas / Marisa Fabiana Haderne ; Compilación de Marisa Fabiana Haderne. - 1a ed. - Mendoza : Universidad Nacional de Cuyo. Secretaría de Investigación, Internacionales y Posgrado, 2025.

Libro digital, PDF

Archivo Digital: descarga y online
ISBN 978-987-575-267-2

1. Computación. 2. Tecnología Informática. 3. Tesis. I. Haderne, Marisa Fabiana, comp. II. Título.

CDD 006.3



Autoridades

Universidad Nacional de Cuyo

Rectora

Cdora. Esther Sanchez

Vicerrector

Mgtr. Gabriel Alejandro Fidel

Facultad de Ingeniería

Decana

Esp. Ing. Patricia Susana INFANTE

Vicedecano

Ing. Mario Carlos SÁNCHEZ

Autoridades Red Unci

Coordinador Titular

Pesado Patricia (UNLP) 2024-2026

Coordinador Alterno

Eterovic Jorge (UNLaM) 2024-2026

Junta Directiva

Carmona Fernanda (UNdeC) 2023-2025

Lasso Marta (UNPA) 2023-2025

Larrea Martin (UNS) 2023-2025

De Vincenzi Marcelo (UAI) 2023-2025

Printista Marcela (UNSL) 2024-2026

Tugnarelli Mónica (UNER) 2024-2026

Panizzi Marisa (UNAHUR) 2024-2026

Brachetta Mariana (UCh) 2024-2026

Miembro Honorario

De Giusti Armando (UNLP)

Secretarías

Secretaría Administrativa: Mariño, Sonia (UNNE)

Prosecretaría Administrativa: Caballero, Sergio (UNaM)

Secretaría Académica: Russo, Claudia (UNNOBA)

Prosecretaría Académica: Haderne, Marisa (UNCUYO)

Secretaría de Ciencia y Técnica: Rodríguez, Nelson (UNSJ)

Prosecretaría de Ciencia y Técnica: Arroyo, Marcelo (UNRC)

Secretaría de Asuntos Reglamentarios: Alfonso, Hugo (UNLaPam)

Secretaría de Vinculación Tecnológica y Profesional: Bursztyn, Andrés (UTN)

Secretaría de Congresos, Publicaciones y Difusión: Thomas, Pablo (UNLP)

Prosecretaría de Difusión: Malbernat, Lucía (CAECE)

Prosecretaría de Congresos WICC: Buckle, Carlos (2025) (UNPSJB)

Prosecretaría de Congresos TEYET: Chaperon, Gabriela (2025) (UM)

Prosecretaría de Congresos CACIC: Fernández, Juan Manuel (2025) (UNLu)

Comité Académico

Universidad	***Representante Titular***	***Universidad***	***Representante Titular***
UBA – Cs. Exactas	Brusco, Pablo Brusco	UN Chaco Austral	Zachman, Patricia
UN La Plata	Pesado, Patricia	UN del Oeste	Massacesi, Alfredo
UN Sur	Larrea, Martin	UN de Cuyo	Haderne, Marisa
UN San Luis	Printista, Marcela	UNdeMardelPlata	Ríos, Carlos
UNCPBA	Aciti, Claudio	UNdeQuilmes	Pari, Denise
UN Comahue	Grosso, Guillermo	UNHurlingham	Panizzi, Marisa,
UN La Matanza	Eterovic, Jorge	UNSAdA	Ramón, Hugo
UN La Pampa	Alfonso, Hugo	UNSAM	Estayno, Marcelo
UNTierra del Fuego	Koremblit, Gabriel	UNdAV	Mayer, Roberto
UN Salta	Gil, Gustavo	UNLZ	Serra, Diego
UN Patagonia Austral	Lasso, Marta	UTN	Bursztyn, Andres
UN SanJuan	Rodriguez, Nelson	UNGS	Fernandez, Enrique
UADER	Mengarelli, Jose Luis	UMorón	Chapperon, Gabriela
UN Patagonia SJB	Buckle, Carlos	UAI	De Vincenzi, Marcelo
UN Entre Ríos	Tugnarelli, Mónica	UBelgrano	Aguilera, Sergio
UN Nordeste	Mariño, Sonia	Ukennedy	Albarracín, Mario

UN Rosario	Casali, Ana	UAdventistadelPlata	Bournissen Juan
UN Misiones	Caballero, Sergio	UCAECE	Malbernat, Lucía
UNNOBA	Russo, Claudia	UPalermo	Alvarez Adriana
UN Chilecito	Carmona, Fernanda	UCARosario	Grieco, Sebastián
UN Lanús	Azcurra, Diego	USalvador	Zanitti, Marcelo
UN Santiago del Estero	Figueroa, Liliana	UAconcagua	Giménez, Rosana
Universidad de la Defensa	Cipriano, Marcelo	UGastónDachary	Ruidías, Hector Javier
UN Litoral	Loyarte, Horacio	UADE	Feijo, Daniel Jose
UN Rio Cuarto	Arroyo, Marcelo	UCEMA	Guglianone, Ariadna
UN Córdoba	Fridlender, Daniel	UAustral	Cosentino, Juan Pablo
UN Jujuy	Herrera Cognetta, Analía	UAtlántida Argentina	Rathmann, Liliana
UN Rio Negro	Lugani, Carlos Fabian	UCALaPlata	Bertone, Rodolfo
UN Villa María	Prato, Laura	ITBA	Bolo, Mario
UN Lujan	Fernandez, Juan Manuel	UChampagnat	Brachetta, Mariana
UN Catamarca	Poliche Maria Valeria	UMendoza	Asensio, Mariela
UN La Rioja	Molina, Miguel	UMET	Quiroga, Flavia Yanina
UNTres de Febrero	Oliveros, Alejandro	Universidad de Chubut	Casco, Julio Cesar
UN Tucumán	Luccioni, Griselda María	Universidad Siglo 21	Conde, Sergio
UNAJ	Morales, Martín		

Coordinadores Áreas

Agentes y Sistemas Inteligentes.

Marcelo Falappa (UNS)
Franco Ronchetti (unlp)
Daniel Pandolfi (UNPA)

Arquitectura, Redes y Sistemas Operativos

Luis Marrone (UNLP)
Orlando Micolini (UNC)
Alejandro Stankevicius (UNS)

Computación Gráfica, Imágenes y Visualización

Martín Larrea (UNS)
María J. Abásolo (UNLP)
Juan Manuel Fernandez (UNLu)

Ingeniería de Software

Pablo Fillottrani (UNS)
Pablo Thomas (UNLP)

Tecnología Informática aplicada en Educación

Alejandra Zangara(UNLP)
Gustavo Gil (UNSa)
Andres Bursztyn (UTN)

Procesamiento de señales y Sistemas de TR

Oscar Bría (INVAP)
Fernando Tinetti (UNLP)
Nelson Rodriguez (UNSJ)

Bases de Datos y Minería de Datos

Laura Lanzarini (UNLP)
Claudia Deco (UNR)
Norma Herrera (UNSL)

Innovación en Educación Informática

Claudia Russo (UNNOBA)
Lucía Malbernat (UCAECE)

Fernanda Carmona (UNdeC)

Innovación en Sistemas de Software

Monica Tugnarelli (UNER)
Marisa Panizzi (UNAHUR)
Guillermo Feierherd (UNTDF)

Procesamiento Distribuido y Paralelo

Marcelo Naiouf (UNLP)
Marcela Printista (UNSL)
Javier Balladini (UNCOMA)

Sonia Mariño (UNNE)

Seguridad Informática

Paula Venosa (UNLP)
Jorge Eterovic (UNLaM)
Mariana Brachetta (UCH)

Tesis de Doctorado

Marcela Printista (UNSL)
Alejandra Cechich (UNCOMA)
Laura De Giusti (UNLP)
Martin Larrea (UNS)
Claudio Aciti (UNCPBA)
Marcelo De Vincenzi (UAI)
Claudia Russo (UNNOBA)
Marisa Panizzi (UNAHUR)
Emmanuel Frati (UNdeC)

Comité Organizador

Dr. Carlos Adrián Catania
Dr. Pablo Daniel Godoy
Mgter. Marisa Haderne
Dra. Ana Carolina Olivera
Dra. Elina Rocío Pacini Naumovich
Dra. Silvia Raichman
Lic Tatiana Sofía Parlanti
Lic. Luciano Robino
Mgter Javier Rosenstein

Índice

Agentes y Sistemas Inteligentes	1
Análisis de textos estructurados mediante el desarrollo de herramientas para generación de material de entrenamiento	2
Aplicación de IA en salud mental: resultados y nuevas líneas de trabajo del grupo Gemis	7
Aplicación de técnicas de inteligencia artificial generativa a los procesos creativos	12
Asesoramiento genético en oncología usando IA	16
Autoescalado de workflows científicos en cloud usando aprendizaje por refuerzo	20
Caracterización vial centrada en el usuario basada en regresión ordinal mediante redes neuronales siamesas	24
Ciencia abierta en repositorios institucionales: framework para interoperabilidad semántica	28
Conversión de estructuras sintácticas mediante IAs, entre lenguaje español y lengua de señas argentina, manteniendo la semántica	33
De datos a decisiones: viabilidad del aprendizaje profundo en consumos médicos en el ámbito de las obras sociales	38
Desarrollo de herramientas inteligentes para la toma de decisiones en la gestión eficiente de recursos y la seguridad informática	42
Desarrollo de técnicas de inteligencia computacional y artificial aplicadas a problemas genómica estructural y funcional	47
Ecoconversa: evolución hacia el aprendizaje por refuerzo para la promoción de los ODS	50
Enfoques inteligentes para movilidad y medioambiente: transporte multimodal y gestión de residuos costeros	54
Entrenamiento de GPT personalizado para telemetría en tiempo real	59
Estrategias y algoritmos de optimización mediante aprendizaje automático aplicadas a sistemas basados en energías renovables	62

Exploración de ROS (robot operating system) para el uso en vehículos no tripulados de superficie	67
Fine-tuning y adaptación de modelos de lenguaje abiertos en infraestructura hpc para aplicaciones de dominio específico	72
Integración de grafos de conocimiento y modelos RAG para la gestión inteligente de datos oceanográficos	77
Inteligencia artificial para la predicción de riesgo cardiovascular	82
Inteligencia computacional y artificial aplicada a ciudades sostenibles y resilientes	87
Machine learning para el análisis metacognitivo de la sincronización sensorio-motora en runners	92
Métodos para el análisis de datos imprecisos: lógica difusa y r-shiny	97
Modelo de conocimiento como servicio (KaaS) basado en ontologías, grafos de conocimiento, IA y gobernanza de datos	102
Pool de tecnologías 4.0 para impulsar organizaciones regionales	106
Programación basada en agentes para la simulación	111
Sistema para análisis y sustitución de abreviaturas en dominios médicos	116
Traducción de lengua de señas con deep learning	120
Una gramática para una lengua en situación de resistencia: el caso del mapuche	124
Uso de inteligencia artificial como asistente del proceso de desarrollo de software. Un caso testigo	129
Visión por computadoras con aprendizaje profundo	133
Arquitectura, redes y sistemas operativos	138
Dockerización de servidores Scada: ciberseguridad industrial	139
Experimentación de automatización de red utilizando tecnologías de programación y equipamiento de red MikroTik	143
Investigación y desarrollo sobre la enseñanza práctica de tecnologías de internet de las cosas	148

Predicción de solicitudes por segundo http para escalado proactivo de kubernetes: un enfoque basado en machine learning y series temporales	153
Proyecto: sistema de comunicaciones para Cube SAT	156
Tecnologías de IoT y visión artificial con aplicaciones medioambientales y robótica	161
Uso de tecnología de blockchain para mantener la integridad de los datos generados por sensores en aplicaciones de IoT	166
Bases de datos y minería de datos	171
Análisis y manejo inteligente de la información	172
Aplicación de algoritmos de detección de Outliers en el área de salud	177
Aplicaciones de bases de datos espaciales y espacio temporales	181
De los datos al conocimiento: construcción de soluciones de analítica e inteligencia empresarial	186
Hibridación de bases de datos para el análisis de datos masivos	190
Indexación, administración y recuperación de datos multimedia masivos	195
Indexación, búsquedas y aplicaciones de bases de datos no estructurados	200
Líneas de investigación complementarias en el marco de proyectos de aplicación de tecnologías de inteligencia artificial	205
Propuesta de validación de datos abiertos y detección de patrones en salud con machine learning: aplicación en el proyecto INCUCAI (donación y trasplante de órganos y tejidos)	210
Tendencias actuales sobre bases de datos no convencionales	215
Visualizando los datos de DIGIBEPE	219
Computación gráfica imágenes y visualización	224
Análisis visual de datos multidimensionales	225
Análisis visual de datos multidimensionales y espacio-temporales asistido por modelos generativos	230
Análisis Visual y Tecnologías Inmersivas para la Enseñanza en Ciencias de los Materiales	235

Aplicación de TIC para el monitoreo y gestión del ganado en áreas extensas	240
Aplicaciones para adultos mayores y personas con discapacidad auditiva	245
Desarrollo de un modelo de procesamiento y análisis digital de imágenes satelitales para detección de cambios. Estudio de caso: detección temprana de deforestación y degradación en la reserva biosfera Yabotí.	250
Estrategias avanzadas para lograr experiencias virtuales más inmersivas y personalizadas	254
sistema automático de gestión de estacionamiento utilizando inteligencia artificial	259
Ingeniería de software	264
Agrupamiento de escenarios con clustering no supervisado	265
Análisis de desempeño de clasificadores de texto para predecir confianza en equipos virtuales de desarrollo de software	270
Análisis de métodos de estimación de puntos de función y fuentes de datos compatibles, para la automatización del proceso de estimar	276
Aplicación de inteligencia artificial generativa en la gestión de requisitos de software	281
Aplicaciones móviles 3d con realidad virtual y realidad aumentada	286
Aspectos de arquitectura y diseño en aplicaciones distribuidas (web, móviles, Iot, etc) con foco en microservicios	290
Ciudades inteligentes sostenibles. Gobernanza digital, IA en ingeniería de software, buenas prácticas y calidad de la IA	295
Creación de un modelo integral de gestión del conocimiento para las pymes argentinas	300
Desarrollo de herramientas para la operacionalización de buenas prácticas de calidad de software	305
Enfoques y tendencias en el desarrollo de aplicaciones para dispositivos móviles	310
Especificación de requisitos de calidad de software en historias de usuario: una propuesta aplicando IA generativa como soporte	316

Explorando el diseño participativo en la creación de software usando recursos de design thinking: adaptación de recursos	321
Framework de desarrollo de aplicaciones distribuidas para dispositivos ubicuos basado en WS-CDL	326
Hacia la gestión de activos en sistemas de big data basado en CBR orientado a servicios	331
Industria 4.0: estudio de herramientas software para la evaluación de la madurez digital de pymes	336
Ingeniería de software empírica: procedimiento para el desarrollo y mantenimiento de colecciones de proyectos de software	341
Interacción por voz afectiva y segura: un enfoque híbrido de PLN, lógica difusa y blockchain	346
Introducción a la computación cuántica y aspectos de ingeniería de software y bases de datos NoSQL para el desarrollo de sistemas de software híbridos	351
Marco tecnológico para el desarrollo de nano satélites	360
Recuperación y reúso de información relevante en comunidades virtuales	365
Reestructuración de escenarios hacia una visión técnica	370
Técnicas de control numérico y actualización de software en HPC aplicadas al problema de ríos y canales abiertos	375
Tecnologías de apoyo a la informática forense	379
Un modelo de calidad con base en la agilidad para alcanzar la transformación digital	384
Verificación y validación de software en la industria 5.0	389
Innovación en educación en informática	393
Cluster cronos mejorado: expansión y optimización	394
Control adaptado para la interacción natural y accesible robótica inclusiva	398
Del aula invertida al modelo tecnopedagógico en la enseñanza del paradigma paralelo	403
El impacto de inteligencia artificial en la definición, desarrollo y evaluación de competencias	408

Estrategias pedagógicas y recursos didácticos para la enseñanza de las ciencias de la computación	413
Implementación de MQTT y App inventor para el control remoto de robots en entornos de aprendizaje interactivo	418
Incorporando el análisis sintáctico LR canónico y el análisis LALR a la herramienta JFlap	423
Intervención estratégica a través de videojuegos	428
Metodologías y recursos para la enseñanza de la programación: Evaluación con rúbricas, evaluación de calidad y el rol de la IA	431
Plan estratégico para el fortalecimiento de las competencias blandas en el currículo del programa Ingeniería en Software, Informática, Ciencias Computacionales y afines	437
Prominf. Su impacto institucional en la UNLaR.	444
Unidades de proyectos de investigación en Sistemas de Información	450
Innovación en sistemas de software	455
Avances en la línea de investigación: motor de aumentación de prompt modelos de IA generativos	456
El valor del uso de los datos aplicado a la estrategia alimentaria	460
Esquema de información para municipio con capacidad de ciudad inteligente sostenible	465
Estrategias para desacoplar funcionalidades y facilitar el desarrollo de software en instituciones públicas: el caso de la Universidad Nacional de Río Negro	469
El valor del uso de los datos aplicado a la estrategia alimentaria	474
Fortaleciendo la conexión entre proyectos de software y el bienestar comunitario	479
GameCore: Un Framework Backend para Videojuegos Autoritativos	484
Implementaciones Tecnológicas en Ciudades Inteligentes	489
Plataforma inteligente de recomendaciones para optimizar la experiencia del usuario	493

Propuesta de Expansión Regional y Blockchain para la Transparencia en Datos Abiertos	497
Sistema de alertas de geofencing para el monitoreo de adultos mayores basados en android y WEAR OS	502
Tremun: Plataforma para entrenamiento cognitivo mejorada	507
Un marco de gobernanza de datos para la transformación digital en la provincia de Río Negro	512
Procesamiento de señales y sistemas de tiempo real	517
Aplicación de planificación de tiempo real heterogénea en sistemas embebidos, IoT y robótica	518
Deep learning aplicado a la identificación de cantos de anuncio de Boana riojana (Amphibia: anura)	523
Desarrollo de un sistema robótico autónomo para la competición de fútbol robótico	528
Monitoreo continuo de Temperatura Hospitalaria Usando sensores IoT	533
Sistemas de tiempo real, embebidos y simulaciones	538
Telemetría para el monitoreo ambiental en la cámara de fructificación de cultivo de hongos en INTA Mendoza	542
Procesamiento distribuido y paralelo	547
Abordajes en modelado y simulación en sistemas de HPC y salud	548
Algoritmos paralelos en computación de altas prestaciones. Fundamentos, construcción y evaluación de rendimiento.	553
Arquitecturas en la nube: diseño, simulación y despliegue	559
Arquitecturas Multiprocesador en HPC: aspectos de portabilidad, eficiencia energética y computación cuántica	565
Gestión distribuida de datos para aplicaciones de IoT	572
Prevención de enfermedades en el cultivo del arroz aplicando inteligencia artificial y computación de alto desempeño	576
Procesamiento en arquitecturas distribuidas: cloud, fog y edge computing	581

Scheduling en computación serverless	588
Redes de cooperación científica internacionales	594
Acciones de colaboración en el consorcio de cloud computing, big data y emerging topics durante 2024 (CCC-BD&ET)	595
Participación en el Programa UNITA GEMINAE	603
RED DAI: curaduría y gestión de datos abiertos de investigación: una mirada desde el aporte de la educación superior latinoamericana a la ciencia abierta	605
UNI-UEAR - alianza universitaria argentina europea para la transformación digital. Avances de ejecución	610
Seguridad Informática	615
Análisis del aporte de la inteligencia artificial a la mejora de la seguridad en ambientes de internet de las cosas	616
Auditoría en Amazon Web Services	621
Framework de automatización de intercepción de tráfico en aplicaciones móviles basado en Frida	626
Honeypots como fuente de inteligencia de ciberamenazas	629
Inteligencia artificial aplicada a la detección de información de identificación personal (PII) en Argentina	634
OTP-Vote: Mejoras en Seguridad y Verificabilidad para el Voto Electrónico	639
Protección proactiva de la identidad digital de usuarios en web 2.0 y web 3.0 mediante inteligencia artificial	644
Uso de Técnicas de Machine Learning para la detección Temprana de Ransomware	649
Tecnología Informática aplicada a la educación	654
Alfadatizando: Enseñanza en nivel secundario de Pensamiento Computacional para Humanidades Digitales	655
Análisis de la movilidad estudiantil. Formulación de casos de estudio basado en datos de becas Progresar y Belgrano	659

Aplicación de laboratorios virtuales en educación de ciencias exactas	664
Aprendizaje personalizado y colaborativo	670
Avances en la elaboración de estrategias para la inclusión de la ciencia de datos en la escuela secundaria	675
Creación de tecnologías digitales para la innovación educativa	680
Deserción en las carreras de Ingeniería: determinación de perfiles de los estudiantes. Estrategias para disminuir la deserción.	685
Diseño de objetos de aprendizaje usando herramientas de Inteligencia artificial.	690
Diseño de objetos de aprendizaje utilizando metodologías activas, que favorezcan la retención de los estudiantes de la Facultad de Ingeniería de la Universidad Nacional de Lomas de Zamora, en el marco de un plan de estudios por competencias.	695
Enseñar aplicando innovaciones tecnológicas de inteligencia artificial generativa	699
Estrategias para la mejora continua del sistema de aseguramiento de la calidad de la UNDEC	704
Estudio de técnicas de inteligencia artificial para la evaluación en forma conjunta de la usabilidad y accesibilidad de las aplicaciones	710
Evaluación de la experiencia del usuario en plataformas de aprendizaje virtual mediante reconocimiento de emociones	716
Evaluación del sistema de gestión de la calidad en el ámbito educativo	720
Factores que impactan en la permanencia y graduación de estudiantes en carreras de sistemas: un estudio en Universidad CAECE	725
Investigación y desarrollo de tecnologías digitales para educación	728
Laboratorio remoto	733
Observatorio de tecnologías aplicadas a la educación (facena-unne)	737
Procesos de apropiación y uso de tecnologías móviles en Río Negro. Primeros pasos	742
Sobre el diseño del arte y mecánicas de un videojuego para la prevención del Abuso Sexual Infantil	746

Trazando el futuro educativo: exploración, afinamiento y aplicación de grandes modelos del lenguaje en la educación superior	750
Un abordaje de la enseñanza de algoritmos evolutivos en la materia inteligencia artificial en la carrera de ISI UTN FRC desde los ambientes de aprendizaje colaborativos	755
Uso de analítica predictiva para anticipar el desempeño académico en materias del área programación	760
Tesis de Doctorado	765
Análisis de trayectorias utilizando técnicas de minería de datos	766
Estudio de Viabilidad de SYCL como modelo de programación unificado para sistemas heterogéneos Basados en GPUS en bioinformática	774
Integración de métodos de descubrimiento de conocimiento embebido en fuentes de información desestructuradas	790
Método de Auditoría Informática basado en sistemas de procesamiento avanzado de datos para minimizar el riesgo de calidad de los resultados	795
Metodología basada en modelos en tiempo de ejecución para la construcción y operación de sistemas autoconscientes de internet de las cosas	801
Métodos computacionales avanzados para evaluar la correlación temporal entre apellidos y ancestría biológica en Argentina	811
Minería de Textos para la extracción de conocimiento en actividades educativas con información proveniente de redes sociales. Una estrategia didáctica.	816
Modelo de proceso para la construcción y mantenimiento de muestras de proyectos software para estudios en la ingeniería del software empírica	824
Posibilidades pedagógicas de los entornos virtuales 3D en el acompañamiento del aprendizaje de personas con discapacidad auditiva	835
Reconstrucción antropométrica 3D de bajo costo basada en procesamiento de imágenes y Deep Learning	849
Técnicas de Inteligencia artificial explicable basadas en una integración de lógica simbólica y no-simbólica	859

· ASI ·

Agentes y Sistemas Inteligentes

ANÁLISIS DE TEXTOS ESTRUCTURADOS MEDIANTE EL DESARROLLO DE HERRAMIENTAS PARA GENERACIÓN DE MATERIAL DE ENTRENAMIENTO

Marina Cardenas, Julio Castillo, Nicolás Hernández, Juan Carlos Cuevas, Nancy Paez
Laboratorio de Investigación de Software/Dpto. Ingeniería en Sistemas de Información/Facultad Regional Córdoba/Universidad Tecnológica Nacional
{ ing.marinacardenas, dr.jotacastillo, damiannicolas05, juancarloscue, nancydelv}@gmail.com

RESUMEN

En este artículo se describen las herramientas desarrolladas dentro del contexto de las líneas de investigación sobre técnicas de análisis de textos y minería de datos mediante la aplicación de algoritmos de aprendizaje automático y análisis de texto estructurado y no estructurado. Se ha establecido un conjunto de criterios que permiten abordar la problemática de la escasez de material de entrenamiento para el análisis de textos y las herramientas utilizadas para el análisis y procesamientos de grandes volúmenes de textos, con el objetivo de simplificar el trabajo necesario para analizar, crear, y modificar material de entrenamiento para sistemas basados en aprendizaje automático, a la vez que permiten ahorrar tiempo, y disminuir errores manuales cuando se trabaja con textos.

Las tres herramientas detalladas en este artículo, han sido desarrolladas dentro de un proyecto de investigación homologado por la Secretaría de Ciencia y Tecnología (SCyT) de la UTN que tiene por objetivo crear herramientas que permitan analizar grandes volúmenes de datos que se encuentran en forma de archivos de textos estructurados o no estructurados, adecuando la información para su utilización en el entrenamiento de sistemas de análisis de texto. Dentro los archivos a analizar se pueden mencionar los archivos de código fuente de lenguajes de programación como Python, Java, C++ y C, con la reciente incorporación de archivos de Word, PDF y TXT.

Palabras Clave: aprendizaje automático, AHP, análisis de texto, extracción de información, corpus, machine learning, redes neuronales artificiales, texto estructurado, texto no estructurado.

CONTEXTO

Este artículo presenta el proyecto homologado por la SCyT de la UTN denominado "Modelo para el procesamiento de textos estructurados Fase 2" (cód. SIECACO0008518), el cual se enmarca en el ámbito de la lingüística computacional y tiene como objetivo principal el desarrollo de herramientas especializadas para el análisis textual de diversas fuentes de información en formato estructurado.

Actualmente, el proyecto se encuentra dentro del grupo de investigación denominado Grupo de Aprendizaje Automático, Lenguajes y Autómatas (GA^2LA) en donde se desarrollan proyectos relacionados con autómatas y lenguajes formales, procesamiento del lenguaje natural, y aprendizaje automático, y en especial proyectos orientados a la aplicación de la inteligencia artificial para resolver problemas de las ciencias sociales. Físicamente, los integrantes del proyecto desarrollan sus actividades en el Laboratorio de Investigación de Software LIS del Dpto. de Ingeniería en

Sistemas de Información de la Universidad Tecnológica Nacional Facultad Regional Córdoba (UTN-FRC).

1. INTRODUCCIÓN

En este artículo se describen herramientas de software que han sido desarrolladas para el estudio e investigación sobre análisis de texto estructurado y el área de minería de datos. Las mismas, han sido construidas para abordar la problemática del material de entrenamiento en los algoritmos basados en aprendizaje por computadora, reduciendo el trabajo manual de clasificación, anotado y selección que suele ser necesario cuando se trabaja con orígenes de datos basados en texto, y brindar soporte en las prácticas de detección de código fuentes similares con distintos enfoques, desde detección de clones de código fuente que afectan al ciclo de vida y mantenimiento del software hasta detección de prácticas de plagio de software.

La creación y análisis del material de entrenamiento necesario para un sistema de análisis de texto estructurado conlleva una labor intensiva y artesanal que puede resultar desafiante para los anotadores humanos [1, 2]. En respuesta a esta complejidad, las herramientas han sido concebidas con la finalidad de facilitar este proceso. Su enfoque semiautomático no solo agiliza la construcción de corpus, sino que también proporciona un valioso apoyo a los anotadores humanos buscando optimizar la generación y análisis del material de entrenamiento, mejorando así la calidad y la eficacia de los sistemas de análisis de texto estructurado basados en aprendizaje supervisado.

Para ello, en este artículo se describen las tres herramientas desarrolladas: el Programa de Mapeo de Datos (PMD), que analiza y manipula información de diversas fuentes para almacenarla en una base de datos SQL Server. Luego, el Banco de Prueba de Algoritmos de Semejanza (BPAS), que utiliza el método de jerarquía analítica (AHP) para comparar algoritmos de similitud de textos estructurados. Finalmente, el Sistema de Detección de Similitudes (SDS), que procesa archivos de código fuente para detectar prácticas de reutilización mediante algoritmos específicos.

Algunas de las aplicaciones del material de entrenamiento producido por estas herramientas pueden ser verse en problemas de extracción de información y minería de datos en textos estructurados [3, 4], tales como búsquedas de implicaciones en textos [5], implicación textual [6, 7], extracción de información [8, 9] o minería de datos [10].

2. LÍNEAS DE INVESTIGACIÓN

Las líneas de investigación de este proyecto se basan en el uso de redes neuronales más precisamente en aprendizaje supervisado, semi-supervisado y machine learning y su aplicación en el análisis de la Lingüística Computacional. Este campo tiene por objetivo lograr la capacidad de reconocer y comprender el lenguaje humano mediante la creación de modelos computacionales. Una de las problemáticas encontradas para esta área de investigación, es la escasez de material para ser utilizado en el aprendizaje supervisado tanto en el entrenamiento como en las pruebas, de esto se deriva la creación y utilización de corpus, que es el estudio empírico de la lengua a partir de los datos que proporcionan ejemplos reales de producciones lingüísticas.

El equipo de investigación también trabaja en el desarrollo de modelos computacionales en general. Uno de los modelos desarrollados permite modelar la ocurrencia de incendios forestales, y forma parte de otro proyecto de investigación que se denomina Modelos para la Predicción de Incendios Forestales, un proyecto homologado por la SCyT de la UTN.

La línea de innovación está relacionada a la integración de los resultados obtenidos de la lingüística de corpus, de la construcción de

modelos y del uso de técnicas de aprendizaje automático, en el marco del grupo GA^{2}LA.

3. RESULTADOS OBTENIDOS/ESPERADOS

El desarrollo del presente proyecto nos ha dado como resultado un conjunto de herramientas que han sido desarrolladas para el estudio e investigación sobre análisis de texto estructurado y el área de minería de datos.

La primera herramienta se denomina Programa de Mapeo de Datos (PMD), que tiene como objetivo permitir el análisis y la manipulación de la información de diferentes fuentes y orígenes de datos y almacenarlos en una estructura estándar en una base de datos relacional SQL Server, permite tomar datos de orígenes de datos estructurados y registrarlos en una nueva base de datos que se encuentra normalizada para facilitar la búsqueda y análisis de textos.

El PMD está desarrollado tanto en versión de aplicación de escritorio (en el lenguaje de programación C#), como en versión web. Ambas versiones permiten almacenar los datos normalizados en una estructura estándar de una base de datos SQL Server facilitando la búsqueda y análisis de textos. En la Figura 1 se puede visualizar la pantalla donde el usuario establece la conexión con las bases de datos involucradas en el proceso de mapeo / transformación de datos desde un origen hacia un destino.



Figura 1. Interfaz principal de mapeo de datos

La segunda herramienta es un Banco de Prueba de Algoritmos de Semejanza (BPAS) que consiste en una aplicación de consola que utiliza el método de jerarquía analítica (AHP) para realizar comparaciones entre diversos algoritmos que realizan evaluación de la similitud de textos estructurados, aunque también pueden aplicarse sobre textos no estructurados. El objetivo de esta herramienta es el de evaluar la capacidad de la detección de similitudes de código de programas en los diferentes niveles establecidos para el método AHP para determinar cuál es el mejor en base a una serie de criterios analizando código fuente escrito en diversos lenguajes de programación, tales como, Java, C, C#, Perl y Python.

La última herramienta mencionada (SDS), permite realizar el procesamiento de archivo de código fuente con diferentes algoritmos que permitan la detección de prácticas de reutilización para determinar aquellos con mayor similitud. Este sistema tiene por objetivo determinar las similitudes de archivos de código fuente de lenguajes como Java, C++, C y Python, entre otros, trabajando a diferentes niveles de granularidad como módulos, funciones y líneas de código [11]. Esta herramienta ha sido desarrollada con orientación web, para que su disponibilidad sea mayor, por lo que su interfaz se creó mediante la utilización de esas tecnologías tal como se muestra en la figura 2. El SDS permite cargar los archivos necesarios para trabajar en la detección de similitudes, en base a los algoritmos implementados por herramientas disponibles en la Web. Asimismo permite la configuración de los parámetros requeridos por cada algoritmo para poder realizar el análisis comparativo, como así también para almacenar el resultado en un nuevo archivo y generar reportes con las salidas de las ejecuciones realizadas, que permitirán al decisor identificar rápidamente un subconjunto de archivos que presentan gran similitud entre sí, dentro de un conjunto grande de archivos.

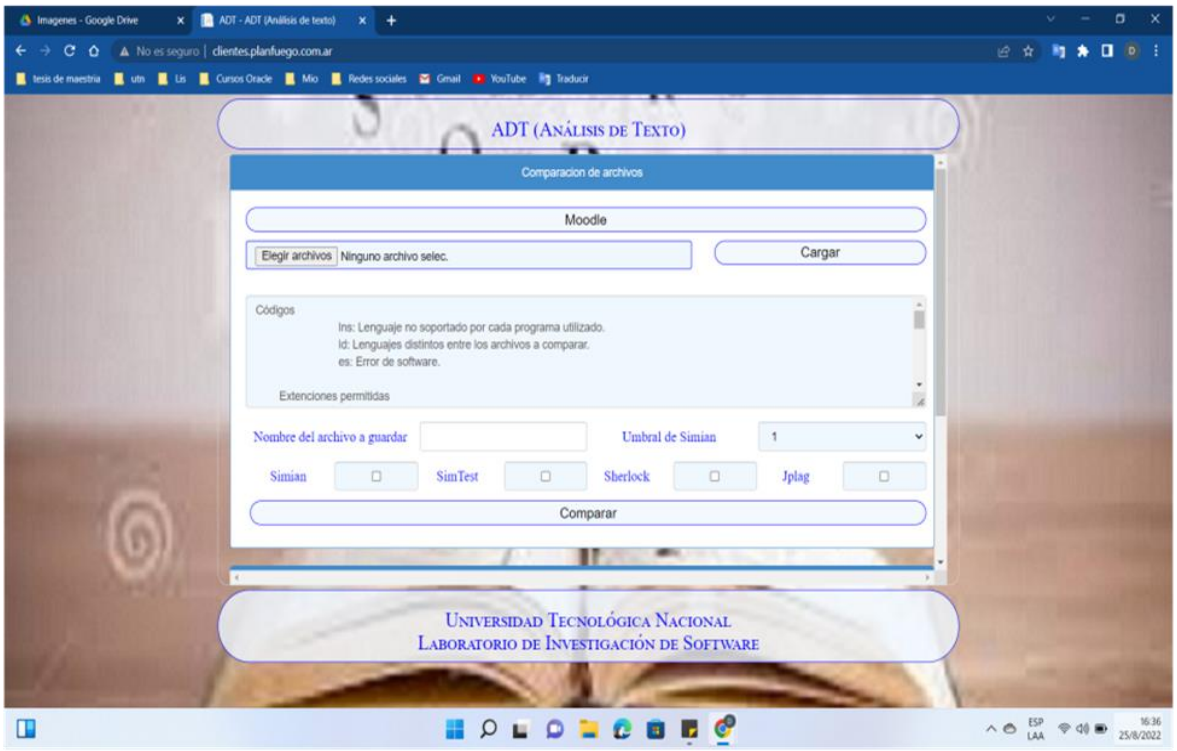


Figura 2. Interfaz principal del SDS.

4. FORMACIÓN DE RECURSOS HUMANOS

- Una doctoranda en ingeniería con mención en sistemas de información en la UTN-FRC, que está trabajando específicamente en el subsistema de detección de similitudes en archivos de código fuente (SDS). Además, realiza la dirección de becarios de posgrado y de becarios de grado en el contexto del proyecto.
- Un doctor en ciencias de la computación quien desarrolló su tesis en el área de investigación y realiza la dirección de becarios de posgrado y de

becarios de grado en el contexto del proyecto.

- Un maestrando en Ingeniería en Sistemas de Información de la UTN-FRC.
- Ocasionalmente participan en el proyecto, alumnos que realizan su práctica supervisada. La cantidad de alumnos depende de las necesidades del proyecto y de los alumnos dispuestos a trabajar en el proyecto. Las prácticas supervisadas son un requisito necesario para la obtención del grado de Ingeniero. Las mismas pueden realizarse en la industria o en el ámbito académico en el contexto de proyectos de investigación homologados.
- Además participan docentes investigadores ingenieros del departamento de Ingeniería en Sistemas de Información de la UTN FRC.
- En el proyecto participan alumnos becarios anualmente, complementando así su formación curricular desde el punto de vista científico.
- Adicionalmente, se realizan charlas de difusión y jornadas de capacitación a alumnos y a docentes de ingeniería en sistemas de información en las líneas temáticas enumeradas anteriormente.

5. BIBLIOGRAFÍA

[1] Castillo, J. y Cardenas, M., "Creación de Corpus para Aplicaciones de Análisis de Texto No Estructurado". CONAIISI 2015. Buenos Aires.

[2] Stefan, T. y Stefanowitsch, A., "Corporain in Cognitive Linguistics", Corpus -Based Approaches to Syntax and Lexis, Berlin: Mouton, 2006, pp. 1-17.

[3] Klavansy, J., Resnik P., "The Balancing Act. Combining Symbolic and Statistical Approaches to Language", MIT Press, 1996.

[4] Manning, C., Schutze H., "Foundations of Statistical Natural Language Processing". The MIT Press, Cambridge, MA, 1999.

[5] Castillo, J., "Sagan in TAC2009: Using Support Vector Machines in Recognizing Textual Entailment and TE Search Pilot task". TAC, NIST, USA, 2009.

[6] Castillo, J. y Cardenas. M., "Using Sentence Semantic Similarity Based on WordNet in Recognizing Textual Entailment". Iberamia 2010, LNCS, vol. 6433, 2010, pp. 366-375.

[7] Castillo, J.,"Using Machine Translation Systems to Expand a Corpus in Textual Entailment", Proceedings of the Icetal 2010, LNCS, vol. 6233, pp. 97-102.

[8] Feldman, R. y Hirsh, H., "Exploiting Background Information in Knowledge Discovery from Text", Journal of Intelligent Information Systems, 1996.

[9] Lewis, D., "Evaluating and optimizing autonomous text classification systems", Proceedings of SIGIR-95, 18th ACM International Conference on Research and Development in Information Retrieval. Seattle, US, 1995, pp. 246-254.

[10] Craven, M., Shavlik, J., "Using Neural Networks for Data Mining", Future Generation Computer Systems, 13, 1997, pp. 211-229.

[11] Castillo, J. y Cardenas, M., Curti, A., Casco, O., Navarro, M., Hernández, N., y Velazco, M., "Desarrollo de sistemas de análisis de texto," in XIX Workshop de Investigadores en Ciencias de la Computación, 2017, pp. 58–62.

APLICACIÓN DE IA EN SALUD MENTAL: RESULTADOS Y NUEVAS LÍNEAS DE TRABAJO DEL GRUPO GEMIS

Di Felice, M.[1]; Lamas, C.[1]; Maleh, F.[1]; Norscini, J.[1]; Ramón, H.[1]; Deroche, A.[1]; Lebedinsky, M[1]; Leguizamón, R[1]; Montenegro Aguilar, G.[1]; Trupkin, I[1]; Pytel, P.[1]; Vegega, C.[1]; Chatterjee, P.[1]; Pollo-Cattaneo, Maria F.[1]

[1] GEMIS - GEMIS.BA
Universidad Tecnológica Nacional. Facultad Regional Buenos Aires.
Medrano 951 (C1179AAQ) Ciudad Autónoma de Argentina. Buenos Aires

flo.pollo@gmail.com

Resumen

La salud mental es un pilar fundamental del bienestar humano, influyendo directamente en la calidad de vida y en el desempeño diario de las personas. En los últimos años, la creciente conciencia sobre la prevalencia y el impacto de los trastornos mentales ha impulsado el desarrollo de enfoques innovadores que complementen las estrategias tradicionales de diagnóstico y tratamiento. En este contexto, la inteligencia artificial (IA) ha emergido como una herramienta transformadora, capaz de procesar grandes volúmenes de datos clínicos, conductuales y genéticos para mejorar la detección temprana de patologías, personalizar intervenciones terapéuticas y optimizar la toma de decisiones en entornos clínicos.

Desde 2022, el Grupo GEMIS-GEMIS.BA de la Universidad Tecnológica Nacional - Facultad Regional Buenos Aires se encuentra trabajando en una nueva línea de investigación sobre la aplicación de IA en salud mental. En esta línea de trabajo, además del desarrollo de modelos predictivos basados en datos clínicos y del análisis automatizado de señales e imágenes para la identificación de nuevos patrones y predictores, se ha incorporado el estudio de los sistemas de apoyo a la decisión clínica *(Clinical Decision Support System - CDSS) basados* en IA. Esta nueva dimensión de investigación busca evaluar la viabilidad, aceptación y potencial de estas herramientas en la práctica clínica, con el objetivo de mejorar la precisión diagnóstica y la personalización del tratamiento. Asimismo, en el marco de este proyecto, se ha ampliado el enfoque metodológico mediante la aplicación de modelos tanto supervisados como no supervisados, lo que permite abordar el desafío de trabajar con datos no etiquetados, una situación común en el ámbito de la salud mental. Esta estrategia busca no solo mejorar la capacidad predictiva de los modelos, sino también ampliar su aplicabilidad en escenarios clínicos reales.

El presente trabajo expone los avances recientes en esta línea de investigación, destacando las contribuciones del uso de IA en salud mental y las oportunidades que ofrecen los enfoques híbridos para el desarrollo de herramientas más precisas y accesibles en el ámbito clínico.

***Palabras clave*:** Inteligencia Artificial, sistemas inteligentes, salud mental, predicción, aprendizaje automático

1. Introducción

La inteligencia artificial (IA) se ha consolidado como una herramienta fundamental en diversas ramas de la medicina, permitiendo la detección temprana de enfermedades, la optimización de tratamientos y el desarrollo de nuevas estrategias terapéuticas. Su capacidad para procesar grandes volúmenes de datos y extraer patrones complejos ha demostrado ser especialmente valiosa en disciplinas como la oftalmología, la oncología y la radiología, donde los modelos de IA han alcanzado, e incluso superado, el rendimiento de especialistas en la identificación de anomalías sutiles en imágenes médicas [1].

En el campo de la salud mental, la incorporación de la IA ha sido más progresiva, dado el carácter altamente individualizado de los trastornos psiquiátricos y la necesidad de enfoques centrados en el paciente. Sin embargo, su potencial para transformar la psiquiatría y la psicología es significativo, particularmente en lo que respecta a la redefinición de los criterios diagnósticos y la caracterización de los trastornos mentales. A través del análisis avanzado de datos, la IA facilita la detección precoz de patologías psiquiátricas, el desarrollo de modelos de riesgo para evaluar predisposiciones individuales y la identificación de patrones subyacentes en el comportamiento y la fisiología de los pacientes. En este contexto, el uso de enfoques computacionales avanzados resulta clave para procesar grandes volúmenes de datos y avanzar hacia una atención en salud mental más personalizada [1]. Si bien existen metodologías estructuradas para el diagnóstico de trastornos como la depresión, la aplicación de la IA no solo agiliza y optimiza este proceso, sino que también permite transformar descripciones subjetivas en métricas objetivas, proporcionando modelos clínicos más precisos. Asimismo, la IA posibilita la identificación de correlaciones entre síntomas que podrían pasar desapercibidas en una evaluación convencional, ofreciendo una visión integral del estado del paciente [2], [3], [4].

Por otro lado, el aprendizaje automático aplicado a la interpretación de bioseñales ha emergido como un enfoque prometedor en este ámbito, facilitando el análisis multimodal de datos mediante técnicas probabilísticas que amplían la comprensión neurofisiológica y conductual de los trastornos mentales [5]. Un área de creciente interés en esta línea de investigación es el desarrollo e implementación de Sistemas de Apoyo a la Decisión Clínica (SADC) basados en IA para la personalización del tratamiento de la depresión. Estos sistemas buscan asistir a los profesionales de la salud mediante recomendaciones fundamentadas en datos clínicos, biomarcadores y modelos predictivos, permitiendo así una optimización en la toma de decisiones terapéuticas.

El estudio de estas tecnologías en el contexto de la salud mental abre nuevas posibilidades para mejorar tanto el diagnóstico como el tratamiento de la depresión y otros trastornos psiquiátricos. Evaluar la viabilidad y efectividad de los SADC basados en IA representa un desafío clave dentro de esta línea de investigación, con el potencial de transformar la práctica clínica y avanzar hacia una atención más precisa, personalizada y centrada en el paciente.

2. Contexto

La Universidad Tecnológica Nacional (UTN) de Argentina ha impulsado la investigación en tecnología e innovación a través de programas especializados, incluyendo la creación del Grupo de Estudio e Investigación en Modelos y Sistemas de Información (GEMIS) en 2009. Formalizado en 2024 como GEMIS.BA dentro de la UTN-FRBA, el grupo se ha consolidado en la modelización de sistemas, inteligencia artificial aplicada e ingeniería de software. Inicialmente centrado en educación e ingeniería, su investigación ha evolucionado hacia la salud, destacándose el proyecto "Inteligencia Artificial para el análisis predictivo en salud mental", iniciado en 2023. Este proyecto busca desarrollar modelos de IA para la detección temprana de trastornos y la optimización de intervenciones clínicas. A partir de 2025, GEMIS.BA ampliará su enfoque hacia la evaluación del impacto y la viabilidad de Sistemas de Apoyo a la Decisión Clínica basados en IA, reforzando su contribución a la investigación aplicada en salud.

3. Objetivos y primeros resultados

El presente trabajo tiene como objetivo principal describir los avances y contribuciones del Grupo GEMIS.BA en la aplicación de técnicas de inteligencia artificial en el ámbito de la salud mental. En particular, la investigación se orienta hacia el desarrollo de modelos predictivos mediante aprendizaje automático supervisado y no supervisado, con el fin de analizar datos clínicos y conductuales para identificar patrones asociados a trastornos mentales [6, 7]. Además, se busca ampliar las capacidades analíticas mediante la integración de señales biomédicas e imágenes, permitiendo la generación de modelos más avanzados y precisos en la detección de patologías psiquiátricas.

Dentro de los objetivos específicos, se destaca la identificación de tareas críticas para la recolección, estructuración y estandarización de datos, asegurando su interoperabilidad con sistemas clínicos existentes. Asimismo, se trabaja en el desarrollo de modelos supervisados para la predicción de trastornos como la depresión, optimizando su desempeño mediante técnicas de ingeniería de características y selección de variables relevantes. En paralelo, se exploran enfoques no supervisados para la identificación de perfiles de riesgo y cohortes con características comunes, facilitando un análisis más holístico de la salud mental. Finalmente, se plantea la implementación de un Sistema de Apoyo a la Decisión Clínica (CDSS) basado en estos modelos, que brinde asistencia a los profesionales de la salud en la evaluación y diagnóstico de trastornos mentales, promoviendo un enfoque más personalizado y eficiente en la atención médica.

En las primeras etapas de esta línea de investigación, se han llevado a cabo diversas revisiones sistemáticas para comprender el panorama global de la aplicación de inteligencia artificial en la salud mental [8]. En particular, se han analizado estudios sobre el uso de modelos de aprendizaje automático en la detección y predicción de trastornos como la depresión, así como investigaciones centradas en los aspectos prácticos de la implementación de CDSS en entornos clínicos. A pesar del crecimiento exponencial de la inteligencia artificial en

salud, estos estudios han identificado múltiples desafíos en la adopción de estas tecnologías en la práctica clínica, incluyendo aspectos regulatorios, interpretabilidad de los modelos y validación en entornos reales.

Paralelamente, se ha avanzado en la construcción y validación de modelos predictivos aplicados a salud mental, utilizando datasets específicos para evaluar su precisión y aplicabilidad en la práctica médica. Inicialmente, se han desarrollado modelos supervisados orientados a la predicción de depresión en poblaciones clínicas, explorando distintas configuraciones algorítmicas y métricas de desempeño. Actualmente, el trabajo se centra en la construcción de un pipeline integral que abarque desde el preprocesamiento de datos hasta la implementación de modelos en entornos clínicos. Este pipeline tiene como objetivo facilitar la integración de herramientas de IA en la toma de decisiones médicas, proporcionando información basada en evidencia para optimizar la evaluación y el tratamiento de pacientes.

Adicionalmente, se ha incorporado el análisis de señales e imágenes como un eje clave para mejorar la capacidad predictiva de los modelos. Este enfoque permite trascender los datos clínicos convencionales, explorando patrones fisiológicos y biomarcadores asociados a trastornos mentales. Mediante técnicas avanzadas de procesamiento de señales y aprendizaje profundo, se busca capturar características sutiles que podrían indicar la presencia de alteraciones psiquiátricas en etapas tempranas. Esto no solo contribuirá a una detección más precisa, sino que también permitirá desarrollar estrategias de intervención más adaptadas a las necesidades individuales de cada paciente.

Estos avances representan un paso significativo hacia la consolidación de herramientas basadas en inteligencia artificial para el apoyo en salud mental, abordando tanto la investigación teórica como la validación práctica en entornos clínicos. A futuro, el enfoque se ampliará hacia el análisis de la viabilidad, aceptabilidad e impacto de los CDSS basados en IA, con el propósito de optimizar su implementación en la atención médica personalizada.

4. Formación de Recursos Humanos

Como parte del proyecto de investigación del Grupo GEMIS.BA sobre la aplicación de inteligencia artificial en salud mental, se ha promovido activamente la formación de recursos humanos, integrando a estudiantes y docentes en distintas etapas de su desarrollo académico y profesional. Este enfoque ha permitido que estudiantes de grado y posgrado participen en investigaciones y tesis orientadas al modelado predictivo mediante técnicas de aprendizaje automático. La reciente incorporación de nuevas líneas de trabajo ha fortalecido esta integración, facilitando la participación interdisciplinaria a través de programas como las Becas de Iniciación en Investigación y Desarrollo. Además de fomentar el crecimiento académico, este enfoque ha permitido la consolidación de un equipo de trabajo con experiencia en la aplicación de modelos de inteligencia artificial en salud mental, contribuyendo a la formación de profesionales con competencias en análisis de datos clínicos, modelado predictivo y desarrollo de sistemas de apoyo a la decisión clínica. La interdisciplinariedad

del equipo ha sido clave para abordar los desafíos de la investigación en este campo, integrando conocimientos provenientes de la informática, la estadística, la psicología y la medicina.

5. Conclusiones

Este trabajo destaca la relevancia de la inteligencia artificial en el análisis y detección de trastornos mentales, subrayando la necesidad de continuar explorando enfoques innovadores para mejorar la atención en salud mental. A lo largo de la investigación, se ha avanzado en el desarrollo de modelos predictivos y sistemas de apoyo a la decisión clínica, con el potencial de optimizar la detección temprana de patologías psiquiátricas y personalizar estrategias de tratamiento. Asimismo, se ha identificado la importancia de fortalecer la integración de herramientas computacionales en el análisis de datos clínicos y de comportamiento, lo que permite una comprensión más profunda de la predisposición a enfermedades mentales. La adopción de enfoques basados en inteligencia artificial abre nuevas oportunidades para la personalización de tratamientos y el desarrollo de intervenciones más efectivas, con el objetivo último de mejorar la calidad de vida de los pacientes.

En este contexto, el Grupo GEMIS continuará trabajando en la validación y aplicación de estos modelos en entornos clínicos reales, abordando los desafíos de implementación y buscando estrategias para su integración en la práctica médica. Además, se mantendrá el énfasis en la formación de recursos humanos, asegurando la preparación de futuras generaciones de investigadores y profesionales en la intersección de la inteligencia artificial y la salud mental. Con estos esfuerzos, se espera contribuir significativamente al avance del conocimiento científico en este campo y al desarrollo de soluciones tecnológicas con impacto positivo en la sociedad.

Referencias

1. Graham, S., Depp, C., Lee, E.E., Nebeker, C., Tu, X., Kim, H.-C., Jeste, D.V.: Artificial Intelligence for Mental Health and Mental Illnesses: an Overview. Curr. Psychiatry Rep. 21, 116 (2019). https://doi.org/10.1007/s11920-019-1094-0.
2. Dozois, D.J.A.: Beck Depression Inventory- II. In: Weiner, I.B. and Craighead, W.E. (eds.) The Corsini Encyclopedia of Psychology. pp. 1–2. Wiley (2010). https://doi.org/10.1002/9780470479216.corpsy0113.
3. Kroenke, K., Strine, T.W., Spitzer, R.L., Williams, J.B.W., Berry, J.T., Mokdad, A.H.: The PHQ-8 as a measure of current depression in the general population. J. Affect. Disord. 114, 163–173 (2009). https://doi.org/10.1016/j.jad.2008.06.026.
4. Kroenke, K., Spitzer, R.L.: The PHQ-9: A New Depression Diagnostic and Severity Measure. Psychiatr. Ann. 32, 509–515 (2002). https://doi.org/10.3928/0048-5713-20020901-06.
5. Sajno, E., Bartolotta, S., Tuena, C., Cipresso, P., Pedroli, E., Riva, G.: Machine learning in biosignals processing for mental health: A narrative review. Front. Psychol. 13, 1066317 (2023). https://doi.org/10.3389/fpsyg.2022.1066317.
6. Di Felice, M., Deroche, A., Trupkin, I., Chatterjee, P., Pollo-Cattaneo, M.F.: Predictive Modeling for Detection of Depression Using Machine Learning. In: Florez, H. and Leon, M. (eds.) Applied Informatics. pp. 47–57. Springer Nature Switzerland, Cham (2024). https://doi.org/10.1007/978-3-031-46813-1_4.
7. Di Felice, M.D., Deroche, A., Trupkin, I., Chatterjee, P., Cattaneo, M.F.P.: Depression and Anxiety Diagnosis Using Unsupervised Learning Approach. In: Florez, H., Galpin, I., and Grévisse, C. (eds.) Joint Proceedings of the ICAI 2023 Workshops WAAI 2023, AIESD 2023, WITS 2023, WKMIT 2023, WSM 2023 co-located with the 6th International Conference on Applied Informatics (ICAI 2023), Guayaquil, Ecuador, October 26-28, 2023. pp. 12–24. CEUR-WS.org (2023)
8. Di Felice M, Trupkin I, Deroche A *et al.* Depression diagnosis using Artificial Intelligence: a systematic review [version 1; peer review: awaiting peer review]. *F1000Research* 2024, 13:1549 (https://doi.org/10.12688/f1000research.158443.1).

Aplicación de Técnicas de Inteligencia Artificial Generativa a los Procesos Creativos.

Carlos A. Catania, Liliana Fiallo, Celeste Ordoñez

Universidad Champagnat Facultad de Informática y Diseño, Sede Central, Mendoza, Argentina.

{cataniacarlos,fiallolilianа}@uch.edu.ar, maria.ordonez@fce.uncu.edu.ar

RESUMEN

La inteligencia artificial generativa ha transformado los procesos creativos al introducir nuevas metodologías en la producción de contenido visual y digital. El presente trabajo explora la aplicación de modelos de Inteligencia Artificial generativa en el diseño gráfico y artes visuales, con el objetivo de desarrollar un sistema de asistencia creativa basado en técnicas de aprendizaje profundo. El estudio se estructura en cuatro etapas principales: análisis del problema, diseño e implementación de un asistente inteligente, experimentación y evaluación. "La evaluación se llevará a cabo desde dos enfoques complementarios: un análisis sistemático con métricas computacionales y una evaluación cualitativa basada en el análisis sensorial de expertos en arte y diseño gráfico. Se espera que los resultados contribuyan al fortalecimiento de la investigación en este campo, al desarrollo de herramientas computacionales para la generación de contenido visual y a la difusión de los hallazgos en ámbitos académicos y profesionales.

CONTEXTO

El presente proyecto se desarrolla en el Instituto de Investigaciones de la Facultad de Informática y Diseño de la Universidad Champagnat (Godoy Cruz, Mendoza), en el marco de la Licenciatura en Sistemas de Información. Este trabajo es parte del proyecto de investigación *Evaluación del impacto de la inteligencia artificial generativa en las prácticas de diseño y arte visual* que dio inicio en septiembre de 2024 en el marco de los proyectos bienales de investigación de la Facultad de Informática y Diseño de la Universidad Champagnat. Es importante destacar que el proyecto cuenta con la colaboración de investigadores de la Facultad de Ingeniería y la Facultad de Artes Visuales y Diseño de la Universidad Nacional de Cuyo.

Palabras Claves: Inteligencia Artificial Generativa, Interacción Hombre Computadora, Asistentes Inteligentes, Creatividad Artística.

1 INTRODUCCIÓN

El desarrollo y aplicación de la inteligencia artificial (IA) generativa [4,6,8] ha redefinido los procesos creativos, estableciendo nuevas dinámicas en la producción de contenido digital [1,7]. Estas tecnologías, fundamentadas en modelos de aprendizaje profundo [3,5], posibilitan la generación automática de imágenes [8], texto [6] y otros formatos, trasladando el enfoque desde la ejecución manual hacia la parametrización, dirección y evaluación de los resultados generados algorítmicamente.

En este contexto, la inteligencia artificial no solo actúa como una herramienta de automatización, sino como un agente colaborativo que amplía las posibilidades

creativas y computacionales. Sin embargo, a medida que estas herramientas impactan en el trabajo creativo, también introducen posibles daños en el ecosistema más amplio de información y medios. A medida que disminuyen el costo y el tiempo necesarios para producir medios sintéticos a gran escala, el ecosistema mediático puede volverse particularmente vulnerable a la suplantación de identidad, la manipulación, la desinformación y la distracción .

Desde la perspectiva de la ciencia de la computación, la inteligencia artificial generativa introduce desafíos y oportunidades en múltiples niveles. A nivel técnico, implica la construcción y entrenamiento de modelos sobre grandes volúmenes de datos, optimizando su capacidad de generalización y coherencia en la generación de contenido [3]. En términos operativos, transforma los flujos de trabajo, donde el usuario ya no interviene únicamente en la producción final, sino en la definición de los parámetros y restricciones que guían al modelo [1,2]. Finalmente, a nivel teórico, plantea cuestiones sobre la creatividad computacional, la validación de los resultados y su impacto en la industria del diseño, la comunicación visual y la producción digital.[1].

Este trabajo explora la aplicación de la inteligencia artificial generativa en los procesos de creación visual y diseño computacional, analizando su impacto desde una perspectiva técnica y conceptual. Se abordarán los métodos de entrenamiento de modelos, la importancia de los conjuntos de datos, la evaluación de resultados mediante métricas computacionales y técnicas cualitativas, así como los desafíos éticos y legales emergentes.

2 LÍNEAS DE INVESTIGACIÓN Y DESARROLLO.

El proyecto se enmarca en dos de las áreas de investigación del instituto de investigaciones de la **Universidad Champagnat**, en particular en la aplicación de técnicas de inteligencia artificial generativa para la generación de asistentes en diversas áreas profesionales

Para el desarrollo del presente proyecto se pueden diferenciar 4 etapas principales:

A) Análisis preliminar del problema. En esta etapa se incluye la recopilación de información bibliográfica sobre la aplicación de técnicas de Inteligencia Artificial Generativa para asistir en los procesos creativos relacionados con las artes visuales y el diseño. Sin embargo, esto no descarta la posibilidad de explorar otras áreas que podrían aplicarse al problema formulado en este proyecto. Durante esta etapa se analizarán los distintos procedimientos necesarios para el desarrollo de asistentes basados en IA generativa y los materiales y métodos comúnmente utilizados.

B) Diseño e implementación de un asistente para la generación de contenidos visuales. Se aplicarán diferentes modelos de inteligencia artificial generativa de uso libre [9] para desarrollar un sistema que asista en los procesos creativos dentro de las áreas de artes visuales y diseño.

El proceso se llevará a cabo en dos etapas. En la primera, se construirá o adaptará uno o varios conjuntos de datos adecuados para la generación de contenidos visuales en una temática específica, priorizando el uso de conjuntos de datos previamente creados y de acceso libre en internet. En la segunda etapa, se utilizará un conjunto de datos especialmente generado a partir de la producción artística y creativa de los miembros del proyecto.

C) Experimentación. En la etapa final del proyecto, se llevará a cabo una evaluación integral del modelo desarrollado. Para ello, se empleará un análisis sistemático basado en métricas computacionales, que permitirá medir el rendimiento del modelo en términos

de precisión, coherencia y calidad de los contenidos generados. Estas métricas incluirán indicadores como la similitud estructural, la fidelidad visual y otros parámetros utilizados en la evaluación de modelos de inteligencia artificial generativa.

Paralelamente, se realizará una evaluación cualitativa mediante técnicas propias del análisis sensorial [10], en la que un grupo de expertos en artes visuales y diseño analizará los resultados generados. A través de encuestas, pruebas de percepción y valoraciones subjetivas, se examinarán aspectos como la creatividad, la originalidad y la relevancia artística de las imágenes producidas. Esta doble perspectiva garantizará una evaluación más completa del impacto y la efectividad del modelo en los procesos creativos.

3 RESULTADOS ESPERADOS

Al término de los dos años de duración del plan de trabajo se pretende que se haya logrado:

1. Fortalecer la línea de investigación en la aplicación de modelos de IA generativa para asistir en los procesos creativos.
2. Contar con un prototipo de una herramienta computacional que permita asistir al proceso de generación de contenido visual de artistas y diseñadores. Esta herramienta deberá generar propuestas alineadas con los lineamientos específicos de cada artista visual o diseñador.
3. Desarrollo de artículos académicos y presentaciones en congresos especializados en arte, diseño y tecnología, que divulguen los hallazgos de la investigación y contribuyan al debate sobre el papel de la inteligencia artificial generativa en la creación artística contemporánea, tanto a nivel local como global.

4 FORMACIÓN DE RECURSOS HUMANOS

Se espera capacitar en el ámbito de la investigación a profesores y alumnos interesados en participar en un entorno académico y tecnológico innovador y a todos aquellos actores interesados en los resultados del proyecto.

Sobre la temática de este proyecto se está desarrollando la Maestría en Arte Latinoamericano de la Lic. Liliana Fiallo, en la Universidad Nacional Cuyo bajo la co-dirección del Dr. Carlos A. Catania.

5 BIBLIOGRAFÍA

1. Epstein, Z., Hertzmann, A., Herman, L., Mahari, R., Frank, M. R., Groh, M., Schroeder, H., Smith, A., Akten, M., Fjeld, J., Farid, H., Leach, N., & Pentland, A. (2023). Art and the science of generative AI: A deeper dive. arXiv preprint arXiv:2306.04141.
2. Jiang, H., Brown, L., Cheng, J., Khan, M., Gupta, A., Workman, D., Hanna, A., Flowers, J., & Gebru, T. (2023). AI Art and its Impact on Artists. En AAAI/ACM Conference on AI, Ethics, and Society (AIES '23) (pp. 12). ACM.
3. Brock, A., Donahue, J., & Simonyan, K. (2019). Large scale GAN training for high fidelity natural image synthesis. International Conference on Learning Representations (ICLR).
4. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877-1901.
5. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. Advances

in neural information processing systems, 27, 2672-2680.
6. Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training. OpenAI.
7. Albar Mansoa, P. J. (2024). *La Inteligencia Artificial de generación de imágenes en arte: ¿Cómo impacta en el futuro del alumnado en Bellas Artes?*. Encuentros, 20, 145-164. https://doi.org/10.5281/zenodo.10052355
8. Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., & Sutskever, I. (2021). *Zero-Shot Text-to-Image Generation*. arXiv:2102.12092. https://doi.org/10.48550/arXiv.2102.12092
9. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). *High-Resolution Image Synthesis with Latent Diffusion Models*. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 10684–10695. https://doi.org/10.1109/CVPR52688.2022.01044
10. Hutmacher, F. (2019). *Why is there so much more research on vision than on any other sensory modality? Frontiers in Psychology, 10*, 2246. https://doi.org/10.3389/fpsyg.2019.02246

Asesoramiento genético en oncología usando IA

Hugo Alfonso[1], Gabriela Minetti[1], Carolina Salto[1], Guillermina Angelucci[2],
Sofía Molina Oderiz[3] y Aldana Torres[3]
[1]Laboratorio de Investigación en Sistemas Inteligentes (LISI)
Facultad de Ingeniería - Universidad Nacional de La Pampa
e-mail: [1]{alfonsoh, minettig, saltoc}@ing.unlpam.edu.ar
[2]Referente provincial de tumores heredofamiliares
Ministerio Salud - Provincia de La Pampa
e-mail: [2]angelucci.guillermina@gmail.com
[3]Facultad de Ingeniería - Universidad Nacional de La Pampa
e-mail: [3]{sofioderiz, aldy.t.31}@gmail.com

Resumen

Los avances en genética han permitido identificar mutaciones asociadas a cáncer hereditario, mejorando el diagnóstico y personalización de tratamientos. En este contexto, los médicos de atención primaria de la salud (APS) juegan un papel clave en la detección temprana de pacientes con riesgo genético. Sin embargo, diversas barreras, como la falta de formación especializada y la dificultad para acceder a guías clínicas actualizadas, pueden retrasar la derivación oportuna a asesoramiento genético.

Esta línea de trabajo busca desarrollar una aplicación móvil basada en inteligencia artificial (IA) para asistir a los médicos de APS en la evaluación del riesgo genético. La herramienta integrará datos clínicos y antecedentes familiares, aplicando protocolos validados para respaldar la toma de decisiones. Un aspecto central del proyecto es la identificación de variables relevantes y la determinación de metodologías de IA que permitan obtener predicciones precisas. Asimismo, se priorizará el desarrollo de una interfaz intuitiva y eficiente, facilitando su uso en la práctica clínica. La implementación de esta aplicación pretende optimizar la identificación de pacientes con mayor probabilidad de portar mutaciones genéticas vinculadas al cáncer, promoviendo intervenciones oportunas basadas en evidencia científica actualizada.

Palabras claves: Inteligencia Artificial, asesoramiento genético oncológico, atención primaria de salud, aplicación médica

CONTEXTO

La forma en que se diagnostican y tratan los cánceres, especialmente los asociados a una predisposición hereditaria, está cambiando radicalmente gracias a los avances en genética en el ámbito de la oncología. Estudios recientes han demostrado que la incorporación de la genética en la oncología permite identificar mutaciones específicas que predisponen a ciertos tipos de cáncer, facilitando diagnósticos más precisos y tratamientos personalizados ([1], [2]). Dado su contacto inicial con el sistema de salud, los médicos de atención primaria de la salud (APS) ocupan una posición estratégica para identificar a aquellos pacientes que podrían beneficiarse de una evaluación genética [3]. La detección temprana de personas en riesgo de cáncer hereditario puede permitir intervenciones preventivas, o un seguimiento más cercano que podría mejorar los resultados de salud [4].

No obstante, existen desafíos importantes que obstaculizan una derivación oportuna, como la falta de conocimiento específico sobre genética y cáncer, la falta de acceso a guías clínicas actualizadas (que incluyan a los síndromes hereditarios más frecuentes) o la confusión sobre cuándo y a quién derivar a los pacientes [5]. La formación limitada en genética entre los médicos de APS puede resultar en subutilización de pruebas genéticas y oportunidades perdidas para la intervención temprana [6]. El desarrollo de una aplicación que brinde a los médicos de APS herramientas claras y basadas en evidencia puede mejorar significati-

vamente la identificación y el manejo de pacientes con riesgo genético.

Esta línea de trabajo está orientada a la creación de una aplicación destinada a los médicos de APS, facilitando la derivación oportuna de pacientes a asesoramiento genético en oncología. La aplicación será desarrollada por estudiantes de la Facultad de Ingeniería de la Universidad Nacional de La Pampa, bajo la supervisión de docentes y en colaboración con personal de asesoramiento genético en oncología y oncología clínica. Particularmente, la Dra. Angelucci es experta universitaria en genética y genómica, referente provincial de tumores heredofamiliares y miembro de la Red Argentina de Cáncer Familiar (RACAF). Esta colaboración interdisciplinaria asegura que la aplicación esté alineada con las necesidades clínicas y las últimas directrices en genética oncológica, promoviendo así un enfoque integrado para el manejo del cáncer hereditario. Es importante destacar que el asesoramiento genético en oncología se encuentra en el tercer nivel de atención médica, dado que es un acto especializado y complejo. Por ello, las consultas requieren la derivación de médicos de primer y segundo nivel. En nuestra provincia, contamos con una sola profesional a cargo de este servicio, lo que hace fundamental una derivación adecuada. A través de este proceso, se obtendrán también los resultados de las encuestas y evaluaciones del prototipo de la aplicación.

1. DESARROLLO

La APS, que incluye médicos generalistas, pediatras, enfermeros y obstetras, tiene ventajas como el fácil acceso a la consulta, conocimiento de la familia, visión global del paciente y seguimiento a través de consultas programadas. Para facilitar el diagnóstico y la atención de enfermedades hereditarias, la APS debe conocer las pruebas genéticas, reconocer las indicaciones de derivación, coordinar el manejo y vigilancia del paciente con el especialista correspondiente y realizar el seguimiento familiar [3].

En la provincia de La Pampa el consultorio de asesoramiento genético en oncología funciona una vez a la semana. Debido a la creciente demanda, los turnos se otorgan con demora de 3 a 4 meses desde la solicitud. Actualmente, la mayoría de las derivaciones provienen de oncólogos y ginecólogos; sin embargo, se busca que los médicos de APS asuman un rol central en la identificación y derivación de personas con antecedentes hereditarios de riesgo. Este cambio permitiría implementar estrategias preventivas en personas sanas con predisposición genética al cáncer.

La medicina actual se está impregnando cada vez más de la genética clínica, la genómica, la biología molecular y la medicina de precisión, elementos que son fundamentales en la práctica habitual de la atención primaria de la salud. Es esencial que los profesionales de APS se familiaricen con los conceptos de genética clínica. La APS es el primer contacto entre los pacientes y los servicios de genética clínica, compartiendo un enfoque biopsicosocial que considera al paciente como parte de una unidad familiar. Este enfoque es útil para orientar a pacientes con enfermedades genéticas o hereditarias [3]. La APS, incluyendo médicos generalistas, pediatras, enfermería y obstetras, tiene ventajas como el fácil acceso a la consulta, conocimiento de la familia, visión global del paciente y seguimiento a través de consultas programadas. Para facilitar el diagnóstico y la atención de enfermedades hereditarias, la APS debe conocer las pruebas genéticas, reconocer las indicaciones de derivación, coordinar el manejo y vigilancia del paciente con el especialista correspondiente y realizar el seguimiento [3].

Para brindar una solución a esta problemática se pretende desarrollar una aplicación móvil, que basada en herramientas de inteligencia artificial (IA)([7], [8], [9]), y teniendo en claro los diversos protocolos existentes en la literatura científica que maneja el equipo de salud, pueda determinar el nivel de riesgo genético de padecer cáncer. Para ello es fundamental realizar el relevamiento de las variables a contemplar para cada paciente, además de identificar las interacciones entre ellas en función de los diversos protocolos existentes a aplicar. La aplicación en desarrollo debe ser flexible, permitiendo que los usuarios, en su mayoría médicos de APS, carguen los datos necesarios y que el algoritmo, basado en protocolos y antecedentes, evalúe el riesgo. Es decir, que permita identificar a los pacientes con mayor probabilidad de portar una mutación genética asociada a cáncer, basado

en los datos clínicos y familiares proporcionados al médico por el paciente. La interfaz de usuario de la aplicación debe ser intuitiva, accesible, fácil de usar y debe optimizar la experiencia del usuario permitiendo a los médicos de APS ingresar los datos del paciente de manera rápida y eficiente. Además, la aplicación debe permitir la actualización e Integración de Guías Clínicas internacionales más aquellas que surjan en genética oncológica, para asegurar que las recomendaciones estén basadas en la evidencia científica actualizada.

El nudo central de la línea de trabajo radica en la determinación de las tecnologías de IA a utilizar para realizar las predicciones con la mayor certidumbre posible. Entre las variantes que se están evaluando podemos mencionar el modelo estadístico Gail [10], el modelo matemático de Tyrer-Cuzick [11], el estudio KARMA[1] (Proyecto de Mamografía del Karolinska para la Predicción del Riesgo de Cáncer de Mama) [12], entre otros.

Modelo Gail

Este modelo estadístico es la base para la herramienta de Evaluación de Riesgo de Cáncer de Mama (BCRAT [2], *Breast Cancer Risk Assessment Tool*). BCRAT utiliza la información personal de una mujer para estimar el riesgo de desarrollar cáncer de mama invasivo durante períodos específicos de tiempo, incluyendo: edad, edad al inicio de la menstruación, edad en el primer nacimiento de un hijo, número de familiares de primer grado (madre, hermanas, hijas) con cáncer de mama, número de biopsias mamarias anteriores (ya sean positivas o negativas), presencia de hiperplasia atípica en una biopsia.

Modelo Tyrer-Cuzick

Este modelo es una herramienta utilizada para evaluar el riesgo de cáncer de mama, que, al igual que el modelo de Gail, incorpora tanto factores familiares como personales para calcular la probabilidad de desarrollar cáncer de mama. Varias aplicaciones se basan en este modelo para evaluar el riesgo de padecer cáncer de mama, entre las más conocidas están: *MagView Tyrer-Cuzick Risk Calculator*[3], *StratifyMDX Tyrer-Cuzick Calculator*[4] y *Gencell Risk Assessment Tool*[5].

Estudio Karma

Este estudio se centra en analizar y determinar los roles e interacciones de diversos factores de riesgo, como la densidad mamográfica, determinantes genéticos y estilo de vida, para desarrollar un modelo integral de predicción de riesgo de cáncer de mama. Este enfoque busca personalizar las estrategias de prevención y detección, adaptándolas al perfil de riesgo individual de cada mujer. En base a este estudio, iCAD desarrolla ProFund AI Risk[6] para la evaluación del riesgo de cáncer de mama, que combina los factores de riesgo utilizados en los otros modelos con características de las mamografías.

2. Resultados Esperados

Se espera que esta iniciativa concluya con el diseño y desarrollo de una aplicación móvil que integre tecnologías de IA para evaluar, con base en antecedentes y protocolos establecidos, la probabilidad de que un paciente porte una mutación genética asociada al cáncer. Ello permitiría derivar al servicio de asesoramiento genético en oncología a los pacientes que realmente lo requieran y, de esta forma, hacer un uso optimizado de ese recurso.

Además, este desarrollo propicia la formación de recursos humanos (RRHH) de varias maneras, a saber:

1. Desarrollo de habilidades técnicas especializadas. Por un lado, las estudiantes de la carrera Ingeniería en Sistemas desarrollarán habilidades avanzadas en programación, desarrollo de aplicaciones móviles, manejo de bases de datos y seguridad informática. Estas habilidades son altamente valoradas en el mercado laboral y aumentan su empleabilidad. Por otro lado, colaborar en el proyecto permitirá a los profesionales de la salud adquirir conocimientos sobre el uso de herramientas tecnológicas en la práctica clínica, mejorando

[1]KARMA: https://karmastudy.org/

[2]BCRAT: https://bcrisktool.cancer.gov/calculator.html

[3]MagView: https://magview.com/ibis-risk-calculator/

[4]StratifyMDX: https://stratifymdx.com/tyrer-cuzick-landing/

[5]GENCELL: https://gencellpharma.com/herramienta-de-estratificacion/

[6]ProFund AI Risk: https://atfisica.com/profound-ai-icad-evaluacion-riesgo-cancer-mama-inteligencia-artificial/

su capacidad para integrar tecnología y atención médica.

2. Fomento de competencias interdisciplinarias. La colaboración entre estudiantes y docentes de la Facultad de Ingeniería con profesionales de la salud fomenta el desarrollo de competencias interdisciplinarias. Los participantes aprenderán a comunicarse y trabajar eficazmente con expertos de otros campos, una habilidad crucial en el entorno laboral moderno.

3. Experiencia en proyectos del mundo real. Participar de un proyecto real, con objetivos claros y resultados tangibles, proporciona una experiencia práctica invaluable. Los estudiantes y profesionales obtendrán una comprensión profunda de cómo gestionar proyectos, resolver problemas y adaptarse a los desafíos del mundo real.

4. Aumento del compromiso y la motivación. Participar de un proyecto con un impacto social significativo puede aumentar el compromiso y la motivación de los estudiantes y profesionales, al ver cómo su trabajo contribuye a mejorar la atención médica y la calidad de vida de la comunidad.

3. Formación de Recursos Humanos

Como se mencionó en el apartado anterior, la formación de recursos humanos es un aspecto importante y con muchas aristas en el desarrollo de esta línea de trabajo. En este marco, se ha promovido la incorporación de estudiantes avanzados de la carrera de Ingeniería en Sistemas de la Facultad de Ingeniería de la UNLPam a la línea de desarrollo. Esta acción les brinda la integración de conocimientos adquiridos en distintas asignaturas a lo largo de su formación. Esta participación no solo fortalece su formación teórica y práctica, sino que también les brinda la oportunidad de aplicar metodologías de investigación en un contexto real. Además, esta experiencia puede servir como base para el desarrollo de su trabajo final de carrera, favoreciendo un proceso formativo integral y vinculado con la práctica profesional. Por otra parte, los docentes-investigadores que integran los proyectos realizan diversos cursos de posgrado relacionados con la temática del proyecto, con el objetivo de sumar los créditos necesarios para cursar carreras de posgrado.

References

[1] A. W. Kurian, K. C. Ward, N. Howlader, D. Deapen, A. S. Hamilton, A. B. Mariotto, D. G. Miller, L. Penberthy, and S. J. Katz, "Genetic testing and results in a population-based cohort of breast cancer patients and ovarian cancer patients." *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, vol. 37 15, pp. 1305–1315, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:106408168

[2] S. Domchek, T. Friebel, C. Singer, D. Gareth Evans, H. Lynch, C. Isaacs, J. Garber, S. Neuhausen, E. Matloff, R. Eeles, G. Pichert, L. Van T'veer, N. Tung, J. Weitzel, F. Couch, W. Rubinstein, P. Ganz, M. Daly, O. Olopade, G. Tomlinson, J. Schildkraut, J. Blum, and T. Rebbeck, "Association of risk-reducing surgery in brca1 or brca2 mutation carriers with cancer risk and mortality," *JAMA*, vol. 304, no. 9, pp. 967–975, Sep. 2010.

[3] I. Ejarque Doménech, P. Marín Reina, S. García-Miñaur Rica, I. Chirivella González, M. T. Martínez Martínez, A. M. García Rodríguez, S. Álvarez de Andrés, and J. J. Tellería Orriols, "Criterios de derivación a genética clínica desde atención primaria. documento de consenso," *Atención Primaria*, vol. 54, no. 12, p. 102501, 2022.

[4] J. L. Berliner, S. A. Cummings, B. Boldt Burnett, and C. N. Ricker, "Risk assessment and genetic counseling for hereditary breast and ovarian cancer syndromes—practice resource of the national society of genetic counselors," *Journal of Genetic Counseling*, vol. 30, no. 2, pp. 342–360, 2021. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/jgc4.1374

[5] J. E. Garber and K. Offit, "Hereditary cancer predisposition syndromes." *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, vol. 23 2, pp. 276–92, 2005. [Online]. Available: https://api.semanticscholar.org/CorpusID:32938090

[6] A. Chou, A. Duncan, G. Hallford, D. Kelley, and L. Dean, "Barriers and strategies to integrate medical genetics and primary care in underserved populations: a scoping review," *Journal of Community Genetics*, vol. 12, 07 2021.

[7] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 4th ed. Pearson, 2020.

[8] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[9] M. Mitchell, *Artificial Intelligence: A Guide for Thinking Humans*. Farrar, Straus and Giroux, 2019.

[10] B. Rockhill, D. Spiegelman, C. Byrne, D. J. Hunter, and G. A. Colditz, "Validation of the gail et al. model of breast cancer risk prediction and implications for chemoprevention," *J Natl Cancer Inst*, vol. 93, no. 5, pp. 358–66, Mar 2001.

[11] J. Tyrer, S. W. Duffy, and J. Cuzick, "A breast cancer prediction model incorporating familial and personal risk factors," *Statistics in Medicine*, vol. 23, no. 7, pp. 1111–1130, 2004.

[12] M. Gabrielson, M. Eriksson, M. Hammarström, B. Gullberg, J. Björk, and K. Czene, "The karma study: A swedish mammography cohort for studying breast cancer risk factors and primary prevention," *Cancer Causes Control*, vol. 28, no. 10, pp. 1149–1156, 2017.

Autoescalado de workflows científicos en Cloud usando Aprendizaje por Refuerzo

L. Robino[a,b]; C. Mateos[b]; E. Pacini[a]; Y. Gari[a]; V. Yannibelli[b]; D. A. Monge[c]

(a) Laboratorio de Sistemas Inteligentes (LabSIn); Facultad de Ingeniería; Universidad Nacional de Cuyo.
(b) Instituto Superior de Ingeniería de Software (ISISTAN); UNICEN - CONICET.
(c) Universidad Nacional de Cuyo
luciano.robino@ingenieria.uncuyo.edu.ar, cristian.mateos@isistan.unicen.edu.ar, epacini@uncu.edu.ar, ygari@uncu.edu.ar, vyannibe@exa.unicen.edu.ar, dmonge@uncu.edu.ar

RESUMEN

Las tecnologías workflow son una herramienta muy útil para desarrollos de software científico y computacionalmente costoso. Por esto, se busca agilizar y abaratar su ejecución. En este contexto, se propone desarrollar estrategias inteligentes de autoescalado en Cloud para lograr estos fines.

En el presente trabajo, se describe una línea de investigación conjunta realizada en los últimos años por el Instituto Superior de Ingeniería de Software (ISISTAN) y el Laboratorio de Sistemas Inteligentes (LabSIn). La línea ha apuntado al desarrollo y mejora de un modelo de Aprendizaje por Refuerzo (AR) en autoescaladores para workflows científicos en Cloud.

Actualmente la mejora de este modelo está enfocada en la combinación de técnicas como Deep Learning, algoritmos evolutivos, Large Language Models (LLM) y comparación de diferentes algoritmos para ahorrar el dinero usado en la ejecución de estos workflow así como en minimizar su tiempo de ejecución.

Palabras claves: Workflow Científico, Aprendizaje por Refuerzo, Cloud Computing

CONTEXTO

Esta línea de investigación se lleva a cabo en el marco de una Tesis de Doctorado en Ciencias de la Computación de la Facultad de Ciencias Exactas de la Universidad Nacional del Centro de la Provincia de Buenos Aires.

Los workflow científicos son una herramienta usada para desarrollar software que desacopla los conocimientos informáticos del *know-how* específico de otra disciplina. La ejecución de workflows científicos en infraestructuras basadas en *Cloud Computing* requiere estrategias inteligentes de autoescalado para su ejecución óptima. Sin embargo, la complejidad del problema de autoescalado lleva a recurrir a técnicas como Aprendizaje por Refuerzo (AR). El uso de AR requiere sucesivas ejecuciones de estos workflows para que el autoescalador aprenda *políticas* eficientes. Por este motivo es necesario contar con técnicas que permitan reducir el costo monetario y el tiempo de ejecución de los workflows durante el aprendizaje.

1. INTRODUCCIÓN

La ciencia desde hace un tiempo ha requerido poder de cómputo para su desarrollo. Antes, se recurría a personas reales para los cálculos; hoy, se recurre a computadoras. Esto ha dado lugar a múltiples herramientas para generar software con fines científicos, por ejemplo los workflows científicos. Los workflows son herramientas que representan problemas complejos en conjuntos de tareas y sus dependencias entre ellas. Estos pueden representarse por medio de grafos acíclicos dirigidos donde las tareas son representadas

por vértices y las dependencias por aristas. El fin de estos workflows es desacoplar dos aspectos. Por un lado, los problemas específicos de un área de la ciencia. Por otro lado, las características y problemas asociados a la ejecución de código en una determinada infraestructura computacional.

Debido a la complejidad de estos workflows, muchas veces la ejecución da lugar a períodos altamente paralelizables o a periodos en donde la ejecución es mucho más secuencial. Esta variación dinámica en las demandas computacionales de los workflows hace necesario el uso de infraestructuras con elasticidad a demanda. Es en este contexto que *Cloud Computing* se presenta como una alternativa para la ejecución de workflows. Sin embargo, el uso de recursos virtualizados en Clouds públicas da lugar a una variabilidad del 20% [1,2] en la ejecución que no puede ser ignorada.

Por lo anterior, para una ejecución óptima de los workflows es necesaria una estrategia inteligente de ejecución dentro de los que se conoce como el problema de autoescalado. El autoescalado es un problema que puede descomponerse en dos partes: (a) el escalado (*scaling*) y (b) la planificación (*scheduling*). Escalado se llama al problema que consiste en decidir si se deben agregar, mantener o reducir el número de recursos en uso. Por otro lado, la planificación consiste en asignar cada uno de estos recursos a las tareas a ser ejecutadas. En todos los casos, se busca optimizar la ejecución usando algún criterio. En esta línea de investigación se busca optimizar el tiempo de ejecución (*makespan*) y el *costo* monetario asociado al alquiler de recursos, balanceando ambos objetivos.

Dada la alta incertidumbre asociada a la alta variabilidad de Cloud y la necesidad de estrategias inteligentes de autoescalado, la técnica conocida como Aprendizaje por Refuerzo (AR) es una candidata ideal para este problema. En AR, un agente interactúa con un ambiente por medio de tres elementos: *estado, acción* y *recompensa* [3]. Los estados codifican información del ambiente que el agente usa para actuar. Las acciones son todas las maneras en que el agente puede hacer que el ambiente pase de un estado a otro. Por último, la recompensa es una señal numérica que el agente busca optimizar. De esta manera, conforme el agente interactúa con el ambiente, aprende qué secuencia de estados y acciones permiten maximizar el valor esperado de recompensa. Este proceso iterativo define una *política*, una función que determina la mejor acción para cada estado.

En problemas como la ejecución de workflows hay estados que definen el inicio y el fin. Por ejemplo, para la ejecución de un workflow, el inicio es el momento en que ninguna tarea ha sido ejecutada aún. Por otro lado, el fin es el momento en que todas las tareas han terminado su ejecución. Estos problemas son episódicos, y una ejecución completa de un workflow constituye un *episodio*. El correcto aprendizaje de una política requiere de la ejecución de múltiples episodios, entre 500 y 1000 [4].

A lo largo del aprendizaje, los algoritmos de AR actualizan los valores esperados de recompensa para pares acción-estado por medio de un proceso iterativo. Estos valores son almacenados en una *función Q-Value*, que dice cual es la recompensa esperada si el agente se encuentra en un determinado estado y toma una determinada acción, de las acciones posibles.

Para actualizar los valores esperados, los algoritmos de AR usan estimaciones anteriores para guiar sus acciones (*bootstrapping*). De

esta manera, la estimación anterior y la diferencia con la nueva observación son combinadas para dar lugar a lo que se conoce como métodos de *diferencias temporales*. Bootstrapping y diferencias temporales son dos ideas claves que le permiten a los algoritmos de AR tener adaptabilidad y poder responder al momento en que el episodio está en ejecución.

La forma en que la recompensa esperada es actualizada, también clasifica los algoritmos de AR entre *on-policy* y *off-policy*. State-Action-Reward-State-Action (SARSA) y Q-Learning son algoritmos on-policy y off-policy respectivamente. Los algoritmos on-policy mantienen la consistencia de la política con respecto a la función Q-Value durante la actualización. Por esto, se dice que estos algoritmos son más conservadores en comparación a los algoritmos off-policy. Los algoritmos on-policy tienden a ahorrar recursos, en comparación con los algoritmos off-policy [3].

La necesidad de múltiples episodios para desarrollar una política óptima enfatiza los problemas de una mala inicialización de la función Q-Value. Como el aprendizaje es un proceso iterativo, si la política inicial es mala, sucesivos intentos llevarán a mayores gastos de tiempo y dinero durante el aprendizaje dado que la política es sub-óptima. A pesar de la relevancia de la inicialización, poco ha sido explorado en torno a cómo inicializar la función Q-Value [5]. Esto se debe a que una buena inicialización de la función Q-Value debe codificar información del entorno adquirida con anterioridad [3]. Es por este motivo que por lo general se opta por inicializaciones sencillas como usar valores constantes. Sin embargo, como se dijo antes, explotar información previa puede llevar a un importante ahorro de recursos.

En este sentido, la creación de estrategias de autoescalado usando AR para workflows científicos se muestra como un área de investigación desafiante donde hay muchas aristas para explorar y mejorar el rendimiento de estas estrategias. Con este fin, se ha desarrollado un modelo para optimizar la ejecución de workflows en Cloud [4]. Este modelo del ambiente para AR ha permitido explorar diferentes facetas que permitan ahorrar tiempo y dinero en la ejecución de workflows.

Un punto final que merece esta área es el tiempo de ejecución típico de los workflows, de días a meses. Considerando estos tiempos de ejecución, el aprendizaje de una política óptima con AR sin inicialización apropiada requeriría de años o décadas. Por esta razón en estos problemas se recurren a simuladores de infraestructuras como CloudSimPlus [6]. De esta manera, la evaluación y experimentación de la calidad de las políticas puede realizarse en horas, en lugar de años.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Conforme a lo anterior, se presentan las líneas de trabajo que están siendo desarrolladas orientadas a reducir tiempo y costo del proceso de aprendizaje de políticas en AR:

- Comparación de algoritmos on-policy y off-policy: Usando el modelo de [4] se busca analizar cuál de los dos tipos de algoritmos tiende a ahorrar el gasto de tiempo y dinero [7].
- ReLIEF (Reinforcement Learning Initialization by Evolutionary Formulation): Utilizando algoritmos evolutivos se desarrolló una política inicial que luego se convirtió en una función Q-Value inicial. A partir de

esto se evalúa cómo afecta el gasto de recursos durante el aprendizaje.

- Inicialización basada en Large Language Models (LLM): Se planea generar funciones Q-Value iniciales usando diferentes LLM para analizar cómo estas nuevas tecnologías afectan el proceso de aprendizaje de políticas.
- Deep Reinforcement Learning: Se planea utilizar el modelo de [4] en conjunto con redes neuronales artificiales como aproximadores universales que representen funciones Q-value. Esto abre la puerta para la exploración de más variantes para el modelo dadas las capacidades de las redes neuronales. De esta manera podría variar tanto la representación de los estados como las acciones posibles a ser usadas.

3. RESULTADOS OBTENIDOS

Actualmente los resultados de la comparación entre Q-Learning y SARSA han sido publicados en [7]. Conforme a lo esperado, SARSA, al ser un algoritmo más conservador, ahorra más tiempo y dinero que su competidor Q-Learning. Adicionalmente, resultados preliminares de ReLIEF muestran que el uso de políticas inicializadas con algoritmos evolutivos permiten que el aprendizaje de políticas de autoescalado utilice menos recursos.

4. FORMACIÓN DE RECURSOS HUMANOS

El primer autor de este trabajo realiza su Tesis Doctoral en el marco de una beca doctoral CONICET, financiada por el Proyecto PUE (Proyecto de Unidad Ejecutora) código 22920170100091CO. La Dra. Garí por otra parte se encuentra usufructuando una beca postdoctoral CONICET. Ambos son dirigidos por investigadores del ISISTAN y el LabSIn de forma conjunta.

5. BIBLIOGRAFÍA

[1] Schad, J., Dittrich, J., Quiané-Ruiz, J.A., 2010. Runtime measurements in the cloud: observing, analyzing, and reducing variance. Proceedings of the VLDB Endowment 3, 460–471

[2] Ericson, J., Mohammadian, M., Santana, F., 2017. Analysis of performance variability in public cloud computing, in: Proceedings of the 2017 IEEE International Conference on Information Reuse and Integration, IEEE. pp. 308–314

[3] Sutton, R.S., Barto, A.G., 2018. Reinforcement Learning: An Introduction. A Bradford Book, USA

[4] Garí, Y., Monge, D.A., Mateos, C., 2022. A q-learning approach for the autoscaling of scientific workflows in the cloud. Future Generation Computer Systems 127, 168–180.

[5] Garí, Y., Monge, D.A., Pacini, E., Mateos, C., García Garino, C., 2021. Reinforcement learning-based application autoscaling in the cloud: A survey. Engineering Applications of Artificial Intelligence 102, 104288.

[6] Silva Filho, M.C., Oliveira, R.L., Monteiro, C.C., Inácio, P.R.M., Freire, M.M., 2017. Cloudsim plus: A cloud computing simulation framework pursuing software engineering principles for improved modularity, extensibility and correctness, in: 2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM), pp. 400–406. doi:10.23919/INM. 2017.7987304.

[7] Garí, Y., Pacini, E., Robino, L., Mateos, C., Monge, D.A., 2024. Online rl-based cloud autoscaling for scientific workflows: Evaluation of q-learning and sarsa. Future Generation Computer Systems 157, 573–586.

CARACTERIZACIÓN VIAL CENTRADA EN EL USUARIO BASADA EN REGRESIÓN ORDINAL MEDIANTE REDES NEURONALES SIAMESAS

Hernán Garrido[a] y Martín Domizio[b]

[a] carloshernangarrido@gmail.com, Facultad de Ingeniería, Universidad Nacional de Cuyo - CONICET

[b] mn.domizio@gmail.com, Facultad de Ingeniería, Universidad Nacional de Cuyo - CONICET

RESUMEN

Este artículo propone un nuevo índice de calidad de carreteras usando técnicas de aprendizaje automático, basado en la vibración experimentada por el usuario. El índice se define como un escalar que, para dos ubicaciones recorridas a la misma velocidad, es mayor en la ubicación de mayor vibración. Se toman registros de vibración geolocalizados en experimentos no controlados con dos pasadas del mismo vehículo a diferentes velocidades. El índice puede traducirse en diagramas de iso-experiencia que son útiles para asistentes de inteligencia artificial o vehículos autónomos para tomar decisiones sobre rutas y velocidades. El índice fue probado en un caso real con un autobús de pasajeros utilizando sensores inerciales y de geolocalización de un smartphone.

CONTEXTO

Existen varios motivos para la creación de un índice de rugosidad de carreteras enfocado en los usuarios. Primeramente, facilitaría la obtención de consentimiento de los usuarios para compartir datos de sensores de sus smartphones (personales) si estos son relevantes para sus intereses. Además, la durabilidad de la suspensión del *golden-car* a una velocidad constante de 80 km/h no es la única información relevante [1]. También podrían darse las siguientes situaciones.

- Flotas de transporte público pueden interesarse en la seguridad y comodidad de los pasajeros de pie.
- Flotas de motocicletas de reparto pueden interesarse en carreteras donde las irregularidades ocasionales pueden evitarse.
- Los autos autónomos necesitan información sobre la comodidad de los ocupantes respecto a las vibraciones.
- Ingenieros de autos de carreras deportivas se interesan en el desplazamiento de las masas de las ruedas.
- Los administradores de trenes están interesados en la probabilidad de descarrilamiento.

Además, un índice de rugosidad para carreteras urbanas debe caracterizar el segmento para las velocidades probables y ser útil para comparar segmentos recorridos a diferentes velocidades.

Las superficies de las carreteras tienen diferentes trayectorias posibles con diferentes rugosidades. Esto es especialmente importante para autos pequeños o vehículos de dos ruedas.

INTRODUCCIÓN

Para evaluar el estado de una carretera, se propone seleccionar una característica de vibración relevante, calculada dentro de una ventana de tiempo, como puede ser la aceleración de vibración vertical RMS dentro del vehículo. Esta característica debe registrarse

en al menos dos pasadas del mismo vehículo sobre un mismo segmento, asegurando que ambas mediciones estén geolocalizadas e incluyan información de velocidad.

El índice resultante permite comparar la vibración entre distintas ubicaciones. Si dos puntos han sido recorridos a la misma velocidad, el índice será mayor en aquel donde la vibración sea más intensa. Cada ubicación tendrá un valor único, a menos que en las dos pasadas el vehículo haya seguido trayectorias diferentes con variaciones en la rugosidad del terreno.

Este índice facilita la toma de decisiones al permitir comparar la calidad de distintos puntos de la carretera, identificar irregularidades en la superficie que el conductor pueda evitar y determinar la velocidad máxima recomendada para una ubicación sin superar un límite de vibración seguro.

DESARROLLO

Se define un modelo de red neuronal diseñado para predecir un índice de rugosidad de la carretera (y_{nn}) a partir de un conjunto de características que incluyen la velocidad del vehículo ($v(t)$) y una medida de vibración ($a(t)$). La característica de vibración seleccionada es la aceleración RMS móvil en el cuerpo del vehículo, lo que permite capturar la experiencia del usuario de manera efectiva.

El modelo está compuesto por una red neuronal totalmente conectada con cuatro capas densas. Idealmente, las etiquetas para y_{nn} podrían ser valores de altura del perfil de la carretera, sin embargo, su recolección es costosa. Para abordar esta limitación, se propone un enfoque de regresión ordinal que permite aprender una versión transformada de $y_{nn} = f(x_{nn}) = H_v(v(t), \lambda)^{-1} a(t)$, donde λ es la longitud de onda del perfil de la carretera, conservando la relación de orden entre las mediciones.

El entrenamiento se realiza utilizando dos redes neuronales siamesas que son copias del modelo $y_{nn} = f(v, a)$. Los datos de vibración y velocidad provienen de un mismo vehículo que atraviesa el camino en dos ocasiones o de dos vehículos similares. Para garantizar la coherencia en la comparación de mediciones, se construyen pares ordenados de datos basados en la correlación temporal dinámica (DTW), buscando igualdad en las salidas de los modelos. Esta igualdad se cuantifica usando como función de salida: $\tanh\big(f(v_{1i}, a_{1i}) - f(v_{2i}, a_{2i})\big)$ y etiquetas nulas. A modo de regularización, para evitar la solución trivial $f(v, a) = 0$, se incluyen en el conjunto de entrenamiento: puntos donde $v_{1i} \approx v_{2j}$ y $a_{1i} > a_{2j}$, con etiquetas 1.0; y puntos donde $v_{1i} \approx v_{2j}$ y $a_{1i} < a_{2j}$, con etiquetas -1.0.

El modelo se implementó en Tensorflow [2] y tanto el preprocesamiento de los datos como el entrenamiento del modelo se realizó en Python.

El índice de rugosidad obtenido es único para cada ubicación del camino. Si dos puntos son recorridos a la misma velocidad, el índice será mayor en aquel donde la aceleración del cuerpo del vehículo sea mayor, reflejando así la intensidad de las irregularidades en la superficie.

RESULTADOS

El vehículo seleccionado fue un autobús urbano de tamaño medio autobús urbano que circula regularmente por Ciudad de Mendoza (Argentina). El trayecto seleccionado fue un segmento de 4.5 km en la ruta 913 del sistema de transporte público; que se muestra en la Fig. 1. Se registraron dos pasadas (GPS y datos de aceleración) en junio de 2024 mediante un smartphone Android de gama media (Motorola Moto G32) con la aplicación Sensor Logger de

Kelvin Choi. El smartphone se colocó firmemente en el asiento de la silla situada más a la izquierda en última fila del autobús.



Figura 1. Los números grises y negros indican el número de muestra. Las intensidades azul y roja de los círculos indican la salida del modelo f para la pasada 1 y la pasada 2, respectivamente.

Los datos del GPS (velocidad, latitud y longitud) estaban disponibles a una frecuencia de muestreo de 1 sps, mientras que los datos de aceleración estaban a 100 sps. Estos últimos se procesaron con una ventana de 5 s para obtener la aceleración RMS móvil a 100 sps; que se submuestreó a 1 sps y se ajustó en el tiempo a los datos de GPS. En la pasada 1, se recogieron 577 muestras; y las muestras de la pasada 2 se emparejaron mediante DTW basado en la ubicación, con las de la pasada 1, como se muestra en la Fig. 2.

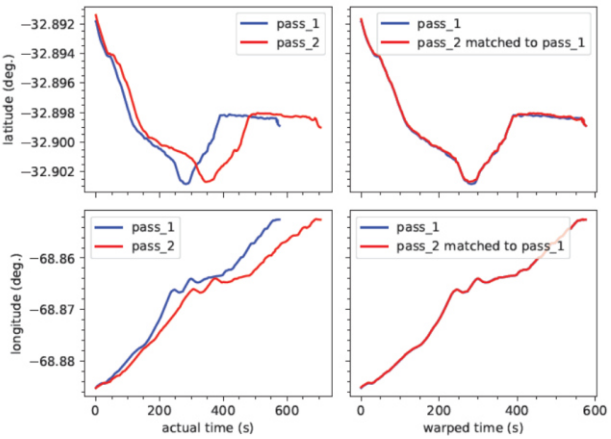


Figura 2. Combado dinámico en el tiempo de las mediciones de velocidad y aceleración.

La figura 3 muestra los resultados de este ejemplo real. (a) velocidades de la pasada 1, y de la pasada 2 combadas para que coincidan con la pasada 1; (b) aceleraciones RMS móvil de las pasadas 1, y 2 combada para que coincidan con las de la pasada 1; (c) resultados de la regresión ordinal.

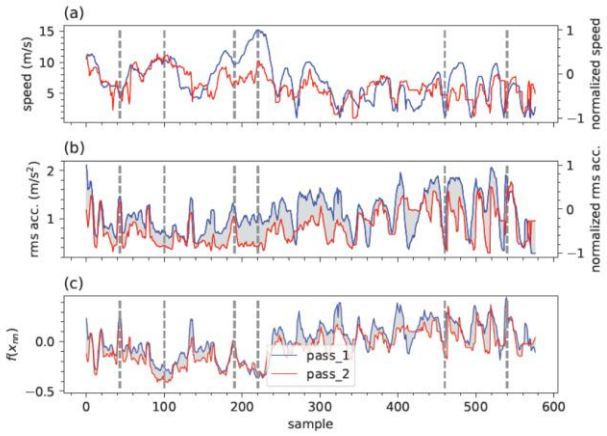


Figura 3. Resultados del entrenamiento.

Las figuras 4 y 5 muestran fotos en correspondencia con los valores obtenidos para el índice de calidad de la carretera. La figura 5 muestra un diagrama de iso-experiencia construido a partir de los datos de la figura 3.

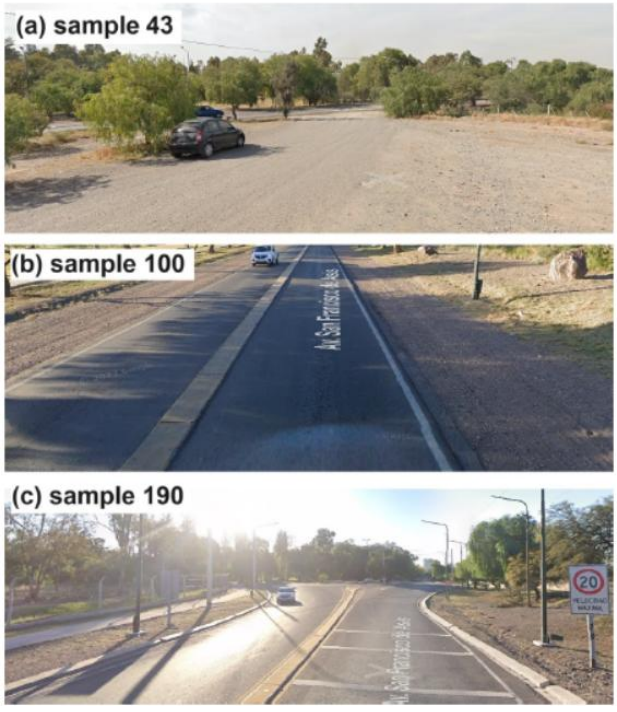


Figura 4. Fotos de la carretera en algunos puntos. (a) Muestra 43: intersección de una carretera interior con la

principal ($y_{nn} = 0{,}23 \pm 0{,}02$). (b) Muestra 100: pavimento de asfalto ($y_{nn} = -0{,}32 \pm 0{,}06$). (c) Muestra 190: Reductores de velocidad sobre pavimento asfáltico ($y_{nn} = -0{,}02 \pm 0{,}03$).



Figura 5. Fotos de la carretera en algunos puntos. (a) Muestra 220: pavimento de asfalto ($y_{nn} = -0{,}35 \pm 0{,}02$). (b) Muestra 460: pavimento de hormigón ($y_{nn} = -0{,}07 \pm 0{,}06$). (c) Muestra 540: pavimento de hormigón dañado ($y_{nn} = 0{,}38 \pm 0{,}02$).

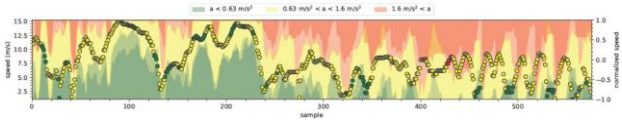


Figura 6. Diagrama de iso-experiencia.

CONCLUSIONES

Este trabajo propone un índice de caracterización de carreteras centrado en el usuario y basado en una magnitud física experimentada por el usuario. La magnitud utilizada en este documento es la aceleración vertical RMS móvil en el interior del vehículo, aunque podrían elegirse otras características de vibración. Se sabe que estas características de vibración dependen en gran medida de la velocidad de desplazamiento, por lo que caracterizar las ubicaciones de la carretera basándose en ellas es difícil cuando la velocidad tiene una alta varianza. Este problema se resuelve definiendo el índice en función de esta aceleración y de la velocidad de desplazamiento.

La metodología fue probada en un caso real utilizando datos obtenidos de un smartphone, con sensores inerciales y de localización, instalado en un autobús de transporte público. Los resultados cuantitativos sugieren que el índice desarrollado es consistente y robusto, y puede ser utilizado para comparar segmentos de carretera en términos de confort y seguridad percibidos.

Por otro lado, se pudieron generar diagramas de iso-experiencia a partir de datos del índice identificado. Estos aportan información sobre el rango de velocidades a las que se puede transitar un determinado camino asegurando cierto nivel de confort y seguridad.

BIBLIOGRAFÍA

[1] Belzowski, B., Ekstrom, A., 2015. Evaluating Roadway Surface Rating Technologies. Technical Report. University of Michigan. Michigan. URL: https://deepblue.lib.umich.edu/bitstream/handle/2027.42/111891/103192.pdf?sequence=1&isAllowed=y

[2] Martin Abadi, 2015. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. Zenodo URL: https://www.tensorflow.org/ , doi:10.5281/zenodo.4724125

CIENCIA ABIERTA EN REPOSITORIOS INSTITUCIONALES: FRAMEWORK PARA INTEROPERABILIDAD SEMÁNTICA

Marcelo P. Amaolo[1] Laura A. Cecchi[1] Germán A. Braun[1] Pablo R. Fillottrani[2 3]

email:{marcelo.amaolo, lcecchi, german.braun}@fi.uncoma.edu.ar, prf@cs.uns.edu.ar

[1] *Grupo de Investigación en Lenguajes e Inteligencia Artificial*
Facultad de Informática
UNIVERSIDAD NACIONAL DEL COMAHUE

[2] *Laboratorio de I&D en Ingeniería de Software y Sistemas de Información*
Departamento de Ciencias e Ingeniería de la Computación
UNIVERSIDAD NACIONAL DEL SUR
[3] *Comisión de Investigaciones Científicas de la provincia de Buenos Aires (CIC)*

RESUMEN

En esta línea de investigación se propone el desarrollo de un framework que guíe la evolución de un Repositorio Institucional (RI) hacia uno que adopte la práctica de Ciencia Abierta, con potencialidad de mejorar su interoperabilidad semántica, y el grado de cumplimiento de principios que aseguren que sus datos y metadatos sean FAIR. Se analiza la representación de los datos y metadatos de RI y RI de Acceso Abierto explorando tecnologías de Datos Enlazados y Web Semántica.

Con la implementación de este framework, se espera que los RI puedan presentar sus contenidos de manera que faciliten el acceso, la disponibilidad y la vinculación con otra información en la red. Esto beneficiará tanto el consumo humano como la recuperación y explotación directa por sistemas informáticos.

Palabras Clave:

Repositorio Institucional - Ciencia Abierta - Interoperabilidad Semántica - Propiedades FAIR - Datos Enlazados.

CONTEXTO

Esta línea de investigación se desarrolla en el contexto de los temas de interés que promueve el Grupo de Investigación en Lenguajes e Inteligencia Artificial (GILIA), de la Facultad de Informática. En particular, se enmarca en el ámbito del proyecto de investigación del GILIA, financiado por la Universidad Nacional del Comahue, *Tecnologías semánticas para el desarrollo de agentes inteligentes (04/F020)*. El proyecto de investigación tiene una duración de cuatro años, a partir de enero de 2022. Asimismo, está financiado por la Universidad Nacional del Sur a través del proyecto de investigación *Tecnologías semánticas para acceso e integración de datos (24/N049)*.

1. INTRODUCCIÓN

Según [BW04], un Repositorio Institucional (RI) es una base de datos compuesta por un grupo de servicios destinados a capturar, almacenar, ordenar, preservar y redistribuir documentación académica en soporte digital. La implementación de un repositorio digital ofrece diferentes beneficios tanto para investigadores y estudiantes, así como al resto de la sociedad, ya que permiten crear y compartir conocimiento, y facilitan la transferencia de conocimiento al sector productivo [War04].

El movimiento de acceso abierto [Bab13], también llamado Ciencia Abierta o directamente Acceso Abierto (Open Access) tiene como objetivo proveer de acceso on-line a toda la información científica disponible de forma gratuita para el lector, para que pueda ser usada y explotada por los investigadores, las empresas y los ciudadanos. La idea que subyace el acceso abierto es que los

resultados y los datos utilizados en investigaciones financiadas por entidades públicas deberían estar disponibles para todos y tener licencias que permitan su reutilización [Cha+02, Bai06].

El desarrollo de *Repositorios Institucionales de Acceso Abierto (RIAA)* ha sido un tema prioritario en las políticas de educación, ciencia y técnica de muchos países. Por ejemplo, en la Argentina, se han desarrollado RIAA en instituciones del ecosistema científico tecnológico nacional en los que se resguardan documentos científicos y/o académicos siguiendo la filosofía del Acceso Abierto[Cha+02]. En nuestro país, desde el año 2013, el Ministerio de Ciencia y Tecnología impulsó los RIAA a través del Sistema Nacional de Repositorios Digitales en Ciencia y Tecnología (SNRD). La Ley N° 26.899 de "Repositorios Digitales Institucionales de Acceso Abierto" [Nac13], establece la obligatoriedad del acceso abierto a la producción científica financiada con fondos públicos a través de los repositorios digitales que las instituciones deberán crear, mantener e integrar al SNRD. Los investigadores deben "*depositar*" en estos repositorios una copia digital de la versión final de las contribuciones aceptadas en publicaciones periódicas o seriadas[NA17].

La norma establece que estos repositorios deberán ser *compatibles con las normas de interoperabilidad adoptadas internacional-mente*, y *garantizarán el libre acceso a sus documentos y datos* a través de Internet u otras tecnologías de información que resulten adecuadas a los efectos, facilitando las condiciones necesarias para la protección de los derechos de la institución y del autor sobre la producción científico-tecnológica.

El paradigma de los *Datos Enlazados* [Hei+14] emerge como un conjunto de buenas prácticas para conectar, compartir y exponer datos y conocimiento, con tecnología que incluye la aplicación de estándares maduros, propuestos por el World Wide Web Consortium (W3C). Su idea principal es establecer relaciones mediante identificadores únicos denominados URIs, entre los datos heterogéneos existentes en la Web, a fin de enriquecerlos semánticamente. Esto permite la *implementación de búsquedas semánticas de información*, diferenciándose de las actuales, en las que sólo se presentan resultados con respecto a una coincidencia sintáctica ingresada. Los servicios de búsqueda y recuperación que permiten el acceso a esta información semántica, logran mejorar sus resultados, aprovechando los elementos relacionados que fueron cargados de forma explícita o inferidos en forma implícita.

La tecnología de datos enlazados *puede ser una oportunidad* para que los RI puedan presentar sus contenidos en una forma que sea más fácilmente accedida, vinculada y reutilizada. Incluso se puede incorporar más información que la representada inicialmente e integrar con otros dominios, aumentando así su expresividad y facilitando la detección y visibilidad. De este modo, la información que se encuentra disponible en RI y RIAA, que se describen hasta ahora con metadatos poco flexibles, se hace accesible para consultas, utilizando herramientas maduras que se han desarrollado en el contexto de la Web Semántica [Zár+19].

Para compartir datos y otros objetos digitales a través de la web, tanto la industria como la academia han establecido requisitos deseables, con el objetivo de ampliar el acceso y asistir a los millones de usuarios ávidos de servicios ágiles con información precisa y de calidad. Estos avances también se pueden incorporar al dominio de los RI para ampliar el cumplimiento de las bondades de la Ciencia Abierta.

Lograr que los *datos y los metadatos sean FAIR* [Wil+16] (Encontrables, Accesibles, Interoperables y Reusables - acrónimo del inglés a partir de los términos Findable, Accessible, Interoperable, Reusable) es un propósito importante para compartir los datos generados en una investigación. Más importante aún es cuando un investigador necesita encontrar información para continuar y mejorar su trabajo. La gestión de datos científicos es esencial en diversos ámbitos de la producción de ciencia y el mundo productivo, por lo que es necesario avanzar en

el establecimiento de políticas, estándares y prácticas que garanticen los principios FAIR *tanto para aplicaciones como para humanos*.
La búsqueda y recuperación de los archivos digitales que representan la producción científica y académica de los RI se realiza a través de motores de búsqueda que consultan *metadatos*, los cuales describen los objetos de datos que se almacenan en los repositorios. Los metadatos se componen de atributos de catalogación de los documentos y otros objetos digitales, y permiten su hallazgo, reconocimiento e identificación, *sin requerir de la apertura y análisis del objeto digital para conocer su contenido*. En la actualidad, muchos de los RI cuentan con metadatos pobremente estandarizados, casi exclusivamente orientados al uso de las personas, lo que genera varios problemas que *limitan la efectividad y utilidad de los mismos*: dificultan la búsqueda precisa de recursos, limitan su visibilidad, generan problemas de interoperabilidad, la evaluación de la calidad y la confiabilidad de la información recuperada. Su mejor utilización requiere del conocimiento y experiencia del usuario que consulta, y se dificulta o imposibilita el tratamiento y entendimiento de las computadoras.
En este trabajo, se propone el desarrollo de un *framework* que guíe la evolución de un Repositorio Institucional hacia uno que adopte la práctica de Ciencia Abierta, con potencialidad de mejorar su interoperabilidad semántica, y el grado de cumplimiento de principios que aseguren que sus datos y metadatos sean FAIR.
El desarrollo del framework se fundamenta en la incorporación de una capa de información semántica, que mapea los datos expresados en los formatos actuales hacia una representación de datos enlazados. Así, se brindan herramientas para solucionar problemas de desambiguación, como los nombres de los autores y se permite incorporar vínculos a otros repositorios y bases de conocimiento publicadas en la web. Además, permite el potencial poblamiento con información aportada por los propios usuarios o inferida a partir de los metadatos y la documentación básica, que pueden ser consultados con herramientas de búsqueda más flexibles mediante el uso del lenguaje de consulta SPARQL.

2. LÍNEA DE INVESTIGACIÓN Y DESARROLLO

En el proyecto de investigación *Tecnologías Semánticas para el desarrollo de Agentes Inteligentes*, UNComa, se investigan técnicas de representación de conocimiento y razonamiento, metodologías de modelado ontológico y mecanismos para la interoperabilidad tanto a nivel de proceso como de datos. El empleo efectivo de los conceptos y conocimientos adquiridos da soporte a comunidades en el desarrollo de ontologías.
Por otro lado, en el proyecto de investigación *Tecnologías semánticas para acceso e integración de datos*, UNS, se desarrollan metodologías y herramientas que asisten la interoperabilidad semántica de información y de servicios en la Web, privilegiando los últimos avances en el área de lenguajes de representación del conocimiento, ontologías y modelado conceptual. Ambos proyectos comparten el perfil de investigación de este trabajo, en el que se estudian entre otros, sobre temas afines a las Tecnologías Semánticas, Interoperabilidad Semántica y Ontologías.
El presente trabajo de investigación se centra en el aprovechamiento de las tecnologías de la Web Semántica, con énfasis particular en aquellas que permiten la implementación de Datos Enlazados, con el objetivo de mejorar la interconexión y la reutilización de los datos y metadatos almacenados. Esta estrategia representa un avance significativo hacia la adopción de los principios FAIR, al establecer un marco de datos y metadatos estandarizado y accesible.
[Umb24] establece cuatro pilares sobre los cuales se basa la nueva construcción y puesta a disposición de los datos productos del quehacer científico como ciencia abierta y establece la importancia de su fusión para aumentar la transparencia, reproducibilidad y calidad de las investigaciones. Del análisis del

estado del arte se observa que existen algunos enfoques que avanzan, por un lado sobre las propiedades necesarias para construir un repositorio abierto [BW04, Hid+19, Raz19], y por el otro para apoyar el proceso de adherir a los principios FAIR [Bah+20, JBA22, Com+21, FI+24, Nos+17, Dil21]. Sin embargo, hasta donde conocemos no hemos encontrado evidencia de que se haya desarrollado a la fecha un *framework* que permita la aproximación simultánea que se propone con nuestra estrategia de trabajo.

3. RESULTADOS OBTENIDOS Y ESPERADOS

Inicialmente, se realizó un relevamiento y análisis de propuestas de representación de la información de las producciones y de los productos digitales generados en el quehacer investigativo presentes en los RI y RIAA. Asimismo, se analizaron las propiedades de la tecnología de datos enlazados y de otros modelos basados en ontologías, y su aplicación en la representación de datos y metadatos de los RI para describir publicaciones científicas y otros objetos digitales.

Con el fin de superar las limitaciones observadas en los Repositorios Institucionales (RI) contemporáneos, se ha desarrollado una primera etapa dentro del marco de trabajo propuesto. Actualmente, nos encontramos realizando un análisis crítico y una categorización de los estándares pertinentes para la representación de datos y metadatos, priorizando la interoperabilidad en un Repositorio Institucional de Acceso Abierto (RIAA). Asimismo, se estudian las propiedades y dimensiones de los principios FAIR para establecer los pasos críticos que faciliten la transformación de datos y metadatos, potenciando así su explotación y aprovechamiento.

Se espera que con el uso de este framework se logre que los RI puedan presentar sus contenidos en una forma que facilite el acceso, la disponibilidad y la vinculación a otra información en la red. Esto favorecerá tanto el consumo humano como la recuperación y explotación directa por sistemas informáticos.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo está conformado por investigadores del Grupo de Investigación en Lenguajes e Inteligencia Artificial de la Universidad Nacional del Comahue y del *Laboratorio de I&D en Ingeniería de Software y Sistemas de Información* de la Universidad Nacional del Sur.

El primer autor está inscripto en la Maestría en Ciencias de la Computación, posgrado acreditado por CONEAU, que dicta la Facultad de Informática de la Universidad Nacional del Comahue, y está delineando su tema de tesis en esta temática.

5. BIBLIOGRAFÍA

[Bab13] Dominique Babini. «UNESCO Guidelines provide a detailed review of Open Access». En: SciELO in Perspective (Blog) (2013) .

[Bah+20] Christophe Bahim, Carlos Casorrán-Amilburu, Makx Dekkers, Edit Herczog, Nicolas Loozen, Konstantinos Repanas, Keith Russell y Shelley Stall. «The FAIR Data Maturity Model: An Approach to Harmonise FAIR Assessments». En: Data Science Journal (oct. de 2020).

[Bai06] Charles W Bailey. «What is open access?» En: Open access: key strategic, technical and economic aspects. Oxford: Chandos Publishing (2006), págs. 13-26.

[BW04] Mary R Barton y Margaret M Waters. «Cómo crear un repositorio institucional». En: Manual LEADIRS II (2004), pág. 169 .

[Cha+02] Leslie Chan, Darius Cuplinskas, Michael Eisen, Fred Friend, Yana Genova, Jean-Claude Guédon, Melissa Hagemann, Stevan Harnad, Rick Johnson, Rima Kupryte et al. Budapest Open Access Initiative. 2002.

[Com+21] European Commission, Directorate-General for Research, Innovation, N Manola, E Lazzeri, M

Barker, I Kuchma, V Gaillard y L Stoy. Digital skills for FAIR and Open Science – Report from the EOSC Executive Board Skills and Training Working Group. Ed. por N Manola, E Lazzeri, M Barker, I Kuchma, V Gaillard y L Stoy. Publications Office, 2021.

[Dil21] Ingrid Dillo. «General overview of FAIRsFAIR». En: (2021) (vid. pág. 7).

[FI+24] GO Fair, GO FAIR Initiative et al. «online».https://www.go-fair.org/go-fair-initiative/.

[Hei+14] Benjamin Heitmann, Richard Cyganiak, Conor Hayes y Stefan Decker. «Architecture of linked data applications». En: Linked Data Management: Principles and Techniques. CRC Press (Taylor & Francis), 2014.

[Hid+19] Yusniel Hidalgo Delgado, Bin Xu, Alejandro Mariño Molerio, Juan Febles-Rodríguez y Amed Mederos. «A Linked Data-based Semantic Interoperability Framework for Digital Libraries». (2019), págs. 14-30.

[JBA22] Clement Jonquet, Syphax Bouazzouni y Emna Amdouni. «O'FAIRe makes you an offer: metadata-based automatic FAIRness assessment for ontologies and semantic resources». En: International Journal of Metadata, Semantics and Ontologies 1 (2022), pág. 1.

[NA17] Silvia Nakano y PA Azrilevich. «El acceso abierto y la implementación de la Ley 26.899 en la Argentina». En: VII Conferencia Internacional sobre Bibliotecas y Repositorios Digitales de América Latina. 2017.

[Nac13] Congreso de la Nación Argentina. Ley 26.899 - Sistema Nacional de Ciencia, Tecnología e Innovación, Repositorios Digitales Institucionales de Acceso Abierto. Boletín Oficial Nro. 32.781 del 09-Dic-2013. Nov. de 2013.

[Nos+17] Brian A Nosek, Jeffrey R Spies, Melanie Benjamin, Natalie Meyers, Sherif Abdelhamid, Dawne Aycock, Luiz Olavo Bonino, Sara Bowman, C Titus Brown, Timothy M Errington et al. «An Open, FAIRified Data Commons: Proposal for NIH Data Commons Pilot». (2017).

[Raz19] Mu Raza. «Application of linked data technologies in digital libraries: a review of literature». En: Library Hi Tech News incorporating Online and CD Notes 36.3 (2019), págs. 9-12

[Umb24] Gaby Umbach. «Open Science and the impact of Open Access, Open Data, and FAIR publishing principles on data-driven academic research: Towards ever more transparent, accessible, and reproducible academic output?» En: Statistical Journal of the IAOS 40.1 (2024), págs. 59-70.

[War04] Mark Ware. «Institutional repositories and scholarly publishing». En: Learned Publishing 17.2 (2004), págs. 115-124.

[Wil+16] Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne et al. «The FAIR Guiding Principles for scientific data management and stewardship». En: Scientific Data 3 (mar. de 2016), págs. 160018-.

[Zár+19] Marcos Daniel Zárate, Carlos Buckle, Renato Mazzanti y Gustavo Samec. «Improving Open Science Using Linked Open Data: CONICET Digital Use Case». En: J. Comput. Sci. Technol. 19.01 (2019), pág. 05.

CONVERSIÓN DE ESTRUCTURAS SINTÁCTICAS MEDIANTE IAs, ENTRE LENGUAJE ESPAÑOL Y LENGUA DE SEÑAS ARGENTINA, MANTENIENDO LA SEMÁNTICA.

Fabrizio Amaya, Gustavo Conturzo, Raúl Oscar Klenzi, María Isabel Masanet

Instituto de Informática / Departamento de Informática /Facultad de Ciencias Exactas Físicas y Naturales / Universidad Nacional de San Juan

Av. Ignacio de la Roza 590 (O), Complejo Universitario "Islas Malvinas", Rivadavia, San Juan, Teléfonos: 4260353, 4260355 Fax 0264-4234980, Sitio Web: http://www.exactas.unsj.edu.ar
e-mail: {fabrizioamaya88, gconturzo, rauloscarklenzi, mimasanet}@gmail.com

RESUMEN

El lenguaje español presenta una estructura sintáctica característica definida por la sucesión de un **sujeto**, la acción ejecutada por él indicada por el **verbo**, continuado por el **objeto** que referencian al sujeto. Una estructura **sujeto - verbo - objeto** (S-V-O).

A su vez, la Lengua de Señas Argentina (LSA) define su estructura sintáctica como **sujeto - objeto - verbo** (S-O-V). Donde el verbo, preferentemente va en infinitivo, ya que el tiempo verbal se expresa con una seña auxiliar que lo indica.

Esta transformación produce un gran estrés mental a las personas intérpretes, dado que se debe esperar que finalice la frase en español SVO, transformarla a SOV y a partir de allí generar la correspondiente secuencia de señas que interprete el mensaje original (interpretación consecutiva), más aún, cuando la interpretación es simultánea, que es la más usada. El advenimiento de diferentes plataformas de Inteligencia Artificial (IA) como ChatGPT, Copilot, DeepSeek, entre otras; basadas en grandes modelos de lenguajes (LLM), ha favorecido un sinnúmero de aplicaciones. Este trabajo pretende, mediante la aplicación de IA, automatizar el proceso de transformación SVO en SOV y viceversa, como parte del proceso de desarrollo de una aplicación web que automatice la comunicación e interacción entre la comunidad sorda y la oyente, habilitando un camino bidireccional entre la LSA y el texto o audio en español.

Palabras clave: Machine Learning, Deep Learning, LLM, LSA.

CONTEXTO

En el marco del proyecto CICITCA-UNSJ "Machine Learning y Deep Learning aplicado a la Lengua de Señas Argentina (LSA)" llevado adelante en el trienio 2023-2025 en dependencias del Departamento de Informática (DI) y del Instituto de Informática (IdeI) de la Facultad de Ciencias Exactas, Físicas y Naturales (FCEFN) dependiente de la Universidad Nacional de San Juan (UNSJ), se intenta construir una aplicación web que automatice la comunicación entre personas de la comunidad sorda y la comunidad oyente.

Si bien tras la pandemia se intentó potenciar las habilidades de las personas en el uso de la LSA, aún es un proceso incipiente. Ante el desconocimiento de la mayoría de la sociedad de la LSA, y la necesidad de que quienes usan esta lengua puedan comunicarse con oyentes, son ejes de trabajo del proyecto.

1. INTRODUCCIÓN

La estructura sintáctica de una oración expresada en formato de texto, extraído de la secuencia de señas en LSA generadas por una

persona sorda, responde al esquema **sujeto - objeto - verbo**. Es necesario realizar una transformación que presente una estructura **sujeto - verbo - objeto,** inherente al formato de texto en lenguaje español para su interpretación por una persona que desconoce la LSA. Este proceso de carácter bidireccional es el que se presenta en la Figura 1 y habrá de ejecutarse mediante aplicaciones basadas en grandes modelos de lenguajes LLM.

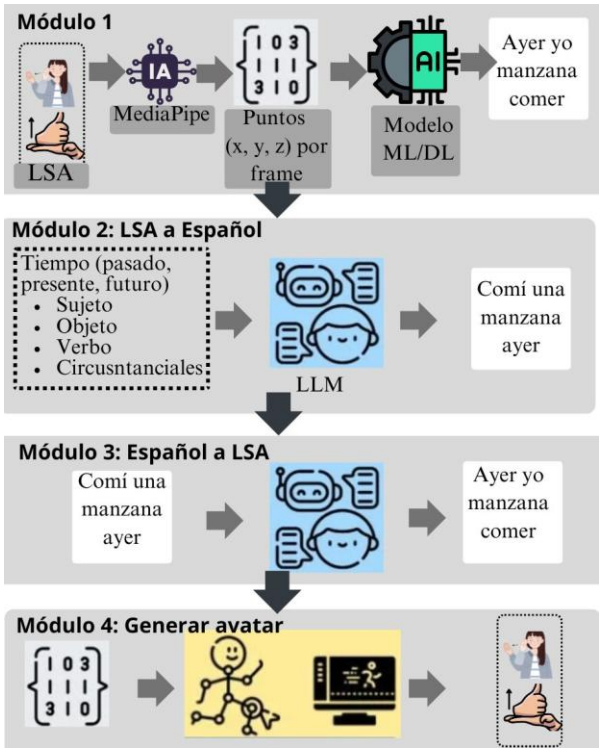


Figura 1: Secuencia de módulos que interpretan la LSA

En la Figura, el módulo 1 representa tareas ya realizadas. Involucra el preprocesamiento de imágenes con herramientas como Media Pipe holistic, los filtrados y muestreos favorecieron la aplicación de algoritmos de Redes Neuronales y XGBoost, y la obtención final del texto asociado a palabras y frases originalmente señadas. Los restantes módulos se están implementando, destacándose el módulo 2 con la conversión de la estructura sintáctica de LSA a español y en módulo 3 la conversión de la estructura sintáctica del lenguaje español hacia la LSA. El módulo 3, es la transformación de la estructura sintáctica del español a LSA mediante un avatar, que simule la secuencia de movimientos asociada al texto obtenido. Los módulos superior derecho e inferior izquierdo son los que se están presentando en su estado de avance en el actual trabajo o propuesta.

La importancia de transformar la estructura sintáctica es crítica en el sentido que, según la semántica que se desprenda de aquella, serán los movimientos de extremidades, corporales y/o gestuales que ejecutará el intérprete o la persona sorda o hipoacúsica.

Es de destacar que en las lenguas escritas se utilizan recursos como metáforas, descripciones y expresiones de lo abstracto; en cambio, las Lenguas de Señas describen lo concreto. La comunidad sorda no entiende lo abstracto, ni metáforas, ni refranes. Así, en el caso de un refrán, el intérprete primero interpreta un español señado y luego da el significado de ese refrán.

Asimismo, en la LSA, la realización de señas direccionales modifica el significado o contexto de una oración. Por ejemplo, el verbo "dar" tiene una seña específica, y dependiendo de la dirección en que se ejecute la seña, indicará la conjugación correspondiente. En el contexto de la frase: *"Jorge da un caramelo a Luis"*, la seña va de izquierda a derecha o de derecha a izquierda dependiendo de la posición de Jorge y Luis respectivamente si estuvieran presentes, en caso de no estarlo, se los ubica espacialmente al indicar el movimiento inicial y final de la seña dar.

Además, como regla general, las negaciones verbales en LSA se construyen agregando la seña correspondiente a la palabra "**no"** después del verbo. A modo de ejemplo, la frase "*dejan de ser aplicables"* en español sería: "*no aplican"* en tanto en LSA sería "*aplicar no"*.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

A lo largo del proyecto, en lo atinente a lograr automatizar el proceso de interpretación de la

LSA, se realizó el procesamiento de videos de diferentes intérpretes y señas en LSA (Franco Ronchetti 64) los cuales fueron tratados mediante herramientas de software como Media Pipe y posteriormente muestreadas a efectos de reducir la cantidad de datos analizados, favoreciendo así la aplicación de algoritmos de aprendizaje como XGBoost. De esta manera, se generaron los textos correspondientes a un conjunto de palabras y frases en LSA. Esto permitió la defensa de un trabajo final de grado en Licenciatura en Ciencias de la Computación durante 2024.

El surgimiento de Chat GPT en noviembre de 2022 habilitó un crecimiento exponencial en la aparición de herramientas de IA que actualmente abarcan la mayoría de las áreas de conocimiento. Como continuación del proyecto, el presente trabajo se centra en el desarrollo de un sistema de traducción bidireccional entre **Lengua de Señas Argentina (LSA) y español**, basado en un enfoque de **agentes inteligentes** y **modelos de lenguaje**. La investigación se articula en tres líneas principales.

2.1. Desarrollo de Agentes Traductores LSA ↔ Español

La traducción entre LSA y español se realiza mediante el desarrollo de agentes que integran modelos de lenguaje existentes (LLMs) utilizando LangChain. Estos agentes permiten la transformación semántica y sintáctica entre ambas lenguas, adaptando las estructuras gramaticales propias de cada una.

Dado que LSA es una lengua visoespacial y no una transcripción directa del español, el sistema incorpora representaciones vectoriales de secuencias de gestos, obtenidas a partir de modelos de visión por computadora, y las combina con modelos de lenguaje preentrenados para la generación y reformulación de texto en español. Este enfoque permite una mejor adaptación de las traducciones sin perder la estructura original del mensaje en LSA.

2.2. Formalización de Reglas Gramaticales y Sintaxis Adaptada

Para garantizar una traducción precisa, el trabajo se ha basado en el desarrollo de reglas sintácticas y gramaticales específicas para la conversión entre LSA y español. Estas reglas han sido elaboradas con el acompañamiento de investigadores intérpretes integrantes del proyecto y validadas a través de experimentos desarrollados con una selección de modelos de lenguaje de gran escala (LLM), entre ellos DeepSeek, LLaMA y Phi-4.

El enfoque adoptado considera las diferencias estructurales entre LSA y el español, incluyendo la adaptación de tiempos verbales, la resolución de ambigüedades semánticas y la correcta asignación de sujetos y predicados en la generación del discurso en español. A través de pruebas iterativas con los modelos LLM seleccionados, se ha optimizado la generación de traducciones que preserven la coherencia y fluidez del mensaje original en LSA, ajustándose a las particularidades gramaticales de cada lengua.

2.3. Aplicación de Metodología Few-Shot Learning

Dado el escaso volumen de datos etiquetados para la traducción automática de LSA, se adopta una metodología **Few-Shot Learning**, basada en modelos de lenguaje capaces de aprender patrones lingüísticos con un número reducido de ejemplos.

Para optimizar la transferencia de conocimiento entre lenguas, se han diseñado **prompts especializados**, los cuales estructuran ejemplos representativos de traducciones LSA-Español con anotaciones detalladas de contexto y reglas sintácticas. A su vez, se evalúan estrategias de **fine-tuning incremental** sobre modelos de lenguaje previamente entrenados, con el fin de mejorar

la capacidad de generalización de los agentes en escenarios reales de comunicación.

3. RESULTADOS OBTENIDOS Y ESPERADOS

El presente trabajo ha evaluado el desempeño de distintos modelos de lenguaje de gran escala (LLMs) en la tarea de traducción de Lengua de Señas Argentina a español y viceversa. Se realizaron pruebas con los modelos DeepSeek, LLaMA y Phi-4, obteniéndose los mejores resultados con Phi-4, el cual mostró mayor coherencia sintáctica y semántica en la conversión de estructuras lingüísticas.

3.1. Traducción de Oraciones Simples

Las pruebas iniciales se realizaron con frases cortas y simples, obteniendo resultados altamente satisfactorios en la reestructuración de la sintaxis de LSA. Ejemplos de traducción incluyen:

Oración en español: El perro está en el jardín.
Traducción en LSA: Perro jardín estar.
Oración en español: Mañana iremos al cine.
Traducción en LSA: Mañana nosotros cine ir.
Oración en español: El libro está en la mesa.
Traducción en LSA: Libro mesa estar.

Los resultados en este conjunto de pruebas indican que las estructuras gramaticales de LSA fueron correctamente respetadas y que el modelo logró preservar el significado de la oración en español.

3.2. Traducción de Textos Largos y Complejos

A medida que se aumentó la complejidad de los textos, se observaron limitaciones en la preservación de la estructura y la información. En textos científicos, literarios y discursivos, la traducción tendió a omitir fragmentos relevantes, lo que sugiere la necesidad de refinar las reglas sintácticas y mejorar el preprocesamiento de la entrada.

Ejemplo de problema detectado en textos extensos:

Texto original: *"Los agujeros negros han sido una de las predicciones más fascinantes de la teoría de la relatividad general de Albert Einstein…"*

Texto en LSA: *"Predicciones fascinantes teoría relatividad general Albert Einstein. Estrella masiva combustible agota colapsa gravedad propia singularidad forma…"*

Aunque el modelo logra extraer los conceptos clave, se detecta una fragmentación excesiva que afecta la fluidez y la comprensión global del texto.

3.3. Acciones y Perspectivas Futuras

A partir de los resultados obtenidos, se plantea la necesidad de una iteración metodológica en la definición de reglas sintácticas, con especial énfasis en:

- Implementación de un sistema de evaluación basado en análisis sintáctico que identifique la correcta asignación de sujeto, predicado y complementos en la traducción.
- Optimización de la segmentación de frases para evitar la pérdida de información en textos largos.
- Ajuste del proceso de preprocesamiento de entrada y salida para mejorar la generación de frases con estructuras más naturales.

El desarrollo de estos agentes representa un avance significativo en la accesibilidad lingüística y en la mejora de los sistemas de comunicación inclusiva. A futuro, se explorará la integración con modelos multimodales, que combinen información de texto, video y señales electromiográficas, con el objetivo de lograr una traducción más natural y fluida. Asimismo, se analizará la viabilidad de extender la metodología a otras variantes de lengua de señas, permitiendo una mayor escalabilidad del sistema propuesto.

4. FORMACIÓN DE RECURSOS HUMANOS

El proyecto está integrado por docentes investigadores responsables de cátedras de las carreras del DI-FCEFyN y que son inherentes a los diferentes pilares que dan soporte al área de conocimiento de la Ciencia de Datos, el Machine Learning y Deep Learning. Se debe destacar que entre los investigadores, dos de ellos son intérpretes de LSA.

Los docentes investigadores fomentan constantemente la participación de estudiantes que integran el espacio e incitan a aquellos que cursan esas cátedras a interiorizarse de la temática e integrarse al proyecto. Tal es así, que integran el proyecto dos estudiantes avanzados de la carrera Licenciatura en Ciencias de la Computación.

Se publicó el capítulo "Determination of Hyperparameters in the Development of a Frost Predictive Model with Data Science" en el libro Computer Science – CACIC 2023 - 29th Argentine Congress of Computer Science, Editorial Springer. 2024.

Se participó en diversos eventos relacionados a la integración de la IA en la salud y en la comunidad sorda. En organizaciones externas al ámbito universitario se expuso en las Primeras Jornadas de Inteligencia Artificial en el Ministerio de Salud bajo el lema "Cuando la Inteligencia Artificial (IA) se hace herramienta Clínica" desarrolladas en el Hospital Rawson de San Juan. Además, en el Conversatorio "Tecnología, Discapacidad y Empleo", organizado por el Instituto Interdisciplinario de Discapacidad del Foro de Abogados de San Juan.

Dentro del ámbito de la Universidad Nacional de San Juan, en las **Jornadas RID del nuevo Cuyo que incluyó** charlas orientadas a personas con discapacidad. Además, en la **2da jornada Universidad y Discapacidad UNSJ. Día internacional de los derechos de las personas con discapacidad.**

Actualmente se encuentra en instancia de evaluación por la Comisión de Licenciatura un trabajo final de la carrera Licenciatura en Ciencias de la Computación, desde donde el grupo publicó en el XII Congreso Nacional de Ingeniería en Informática/Sistemas de Información (CONAIISI - 2024).

5. BIBLIOGRAFÍA

[1] P. Singh, Ed., *Fundamentals and methods of machine and deep learning : algorithms, tools and applications*, 1st ed. USA, 2022.

[2] S. Dash, S. K. Pani, J. Rodrigues, and B. Majhi, Eds., *Deep learning, machine learning and IoT in biomedical and health informatics : techniques and applications*, Fourth. Massachusetts, 2022.

[3] I. El Naqa and M. J. Murphy, "What Is Machine Learning?," *Mach. Learn. Radiat. Oncol.*, pp. 3–11, 2015, doi: 10.1007/978-3-319-18305-3_1.

[4] Valentina Alto. Building LLM Apps Create Intelligent Apps and Agents with Large Language Models. Packt 2024.

[5] "Señario de términos y expresiones en Lengua de Señas Argentina – CAS." https://cas.org.ar/senario-de-terminos-y-expresiones-en-lengua-de-senas-argentina/ (accessed Dec. 03, 2022).

[6] M. I. Massone, "Lenguas de señas: 'cada comunidad desarrolló la propia por necesidad' | CONICET," 2012. Accessed: Nov. 11, 2022. [Online]. Available: https://www.conicet.gov.ar/lenguas-de-senas-cada-comunidad-desarrollo-la-propia-por-necesidad/.

[7] F. Ronchetti, "Reconocimiento de gestos dinámicos y su aplicación al lenguaje de señas," Mar. 2017, doi: 10.35537/10915/59330.

De datos a decisiones: Viabilidad del Aprendizaje profundo en consumos médicos en el ámbito de las obras sociales

Adriano M. Lusso[1] Lara V. Acuña Bravo[2] Sandra E. Roger[1]

{adriano.lusso, lara.acuna}@est.fi.uncoma.edu.ar
roger@fi.uncoma.edu.ar

[1]*GILIA* [2] *GIISCO*
Departamento de Teoría de la Computación
Facultad de Informática
UNIVERSIDAD NACIONAL DEL COMAHUE

Resumen

El Servicio de Obra Social de la Universidad Nacional del Comahue (SOSUNC) dispone de una considerable cantidad de datos relativos a los consumos médicos de sus afiliados. Este trabajo se enfoca en la aplicación de técnicas de reglas de asociación para la identificación de patrones en estos datos, que abarcan información sobre prácticas médicas, prescripciones de medicamentos y fechas de consumo. El objetivo principal es la creación de un corpus de datos que permita discernir las interrelaciones entre las patologías y los consumos asociados, con el fin de optimizar la toma de decisiones en la gestión de las coberturas médicas.

La metodología propuesta incluye la integración con técnicas de *Deep Learning*, con el propósito de desarrollar modelos predictivos capaces de capturar relaciones complejas y no lineales presentes en los datos. Se busca que estos modelos ofrezcan una mayor capacidad de discriminación y una comprensión más profunda de los patrones de consumo. Por último, se busca proporcionar una herramienta que facilite a SOSUNC la anticipación de futuras necesidades y la optimización de la asignación de recursos, logrando así mejorar la calidad de la atención médica y la gestión integral de los recursos disponibles.

Palabras Clave: Minería de datos, Patrones, Asociación, Aprendizaje Automático

Contexto

Este trabajo está parcialmente financiado por la UNCo, en el marco del nuevo proyecto de investigación *Tecnologías Semánticas para el desarrollo de Agentes Inteligentes*. El proyecto de investigación tiene una duración de cuatro años. Ha comenzado en 2022 y se desarrolla en forma colaborativa con docentes-investigadores de la UNS.

1. Introducción

En la era digital, los sistemas de salud generan grandes volúmenes de datos [1] que, si se analizan adecuadamente, pueden proporcionar hallazgos valiosos. Sin embargo, sin herramientas avanzadas, estos datos suelen no aprovecharse. Las obras sociales, como SOSUNC, enfrentan el desafío de gestionar recursos financieros y médicos limitados. Un análisis detallado de los consumos médicos permite identificar áreas de sobreutilización, subutilización o mal uso de los recursos, lo que puede traducirse en ahorros significativos. Al entender los patrones de consumo de los afiliados, es posible diseñar planes de cobertura más personalizados y eficientes, lo que

mejora la satisfacción del usuario y la calidad de la atención.

Los métodos tradicionales suelen basarse en análisis manuales o herramientas básicas de gestión de datos, lo que limita la capacidad para identificar patrones complejos o tendencias ocultas. Los enfoques tradicionales no suelen incorporar técnicas predictivas, lo que dificulta anticipar necesidades futuras o planificar de manera proactiva. Sin técnicas avanzadas, es difícil identificar grupos de afiliados con necesidades específicas, lo que puede llevar a planes de cobertura genéricos y poco efectivos.

Mediante técnicas de minería de datos y Deep Learning, esta línea de investigación permitirá descubrir relaciones no evidentes entre patologías y consumos médicos, lo que no sería posible con métodos tradicionales. Al proporcionar información basada en datos, permitirá a SOSUNC tomar decisiones más informadas y estratégicas, optimizando la asignación de recursos y mejorando la eficiencia operativa. La aplicación de modelos predictivos permitirá anticipar tendencias en los consumos médicos, lo que facilitará la planificación de recursos y la prevención de problemas de salud en la población afiliada [2, 3, 4, 5]. Al identificar patrones de consumo ineficientes o redundantes, se podrán implementar medidas correctivas que reduzcan costos sin comprometer la calidad de la atención.

El desarrollo del plan de trabajo se realizará en el marco del proyecto de investigación "Tecnologías Semánticas para el desarrollo de Agentes Inteligentes". En dicho proyecto de investigación se desarrolla una línea de investigación que explora sobre temas afines tanto al análisis y desarrollo de técnicas de aprendizaje automático con el objetivo de dar soporte a los agentes inteligentes. Particularmente, se ha escogido experimentar sobre herramientas de minería de datos y deep learning que den soporte a la búsqueda de información en el ámbito de la salud.

2. Línea de Investigación y Desarrollo

El proyecto de investigación *Tecnologías Semánticas para el Desarrollo de Agentes Inteligentes* tiene como objetivo general generar conocimiento especializado en el área de agentes inteligentes capaces de acceder, procesar y recuperar información mediante el uso de tecnologías semánticas. En este sentido, se ha desarrollado una línea de investigación enfocada en el estudio de técnicas inteligentes aplicadas a la gestión y análisis de datos en contextos específicos, como el ámbito de la salud y las obras sociales.

En particular, esta línea de investigación se centra en la aplicación de técnicas novedosas de *Deep Learning* y metodologías de minería de datos para generar modelos predictivos y descriptivos que permitan analizar y comprender los patrones de consumo médico de los afiliados, probando y contrastando las técnicas aplicadas sobre los datos en distintos rangos de tiempo. El propósito de este enfoque es contribuir a la toma de decisiones más informadas, mejorando tanto la calidad de la atención médica como la gestión integral de recursos en el ámbito de las obras sociales.

A través de la exploración de técnicas no supervisadas para segmentar a los afiliados en grupos homogéneos basados en sus patrones de consumo médico. Las mismas serán combinadas con técnicas de explicabilidad de modelos predictivos y descriptivos [6, 7, 8, 9]. Este enfoque resulta esencial en un contexto tan delicado como el de la salud, donde la transparencia y la justificación de las decisiones son cruciales para mantener la confianza y la eficacia en la gestión de las coberturas médicas. La segmentación de afiliados permitirá asignar recursos de manera más eficiente, priorizando a aquellos grupos con mayores necesidades médicas. Además, al comprender en profundidad las necesidades específicas de cada segmento, se podrán diseñar políticas que fomenten la equidad en el acceso a los servicios médicos, asegurando que todos los afiliados reciban una atención adecuada y oportuna.

El objetivo final de este trabajo es la construcción de un corpus de datos reales que sirva como base para la aplicación de las técnicas propues-

tas. Este corpus permitirá a SOSUNC identificar patrones de consumo, anticipar necesidades futuras y optimizar la distribución de recursos. La integración de técnicas de Deep Learning en este proceso posibilita el desarrollo de modelos más sofisticados, capaces de capturar relaciones complejas y no lineales entre los datos, lo que incrementaría la precisión y relevancia de los resultados obtenidos.

En esencia, el propósito de esta investigación es demostrar cómo la aplicación de técnicas avanzadas de Deep Learning y minería de datos puede transformar la gestión de consumos médicos en el ámbito de las obras sociales. Mediante la identificación de asociaciones entre patologías y consumos, así como la segmentación explicativa de los afiliados, se busca proporcionar herramientas que permitan a SOSUNC tomar decisiones más informadas, eficientes y fundamentadas. Esto contribuirá no solo a mejorar la calidad de la atención médica, sino también a optimizar la gestión integral de los recursos disponibles.

3. Resultados Obtenidos y Trabajos Futuros

Se pretende estudiar y analizar diversos enfoques y arquitecturas disponibles en la investigación actual. A partir de este análisis, se busca diseñar e implementar una arquitectura que integre técnicas novedosas y actuales para identificar patrones y relaciones en los consumos médicos de los afiliados de la obra social SOSUNC. En este contexto, se planea desarrollar diferentes artefactos de software utilizando diversas técnicas y metodologías, así como diseñar e implementar nuevas técnicas explicativas para Machine Learning que puedan ser incorporadas en el prototipo.

Además, se propone la creación de un dataset que mejore la representación de los historiales médicos, abordando el problema de la alta dimensionalidad de los estados en un historial médico. Esto implica explorar otras representaciones que permitan mitigar esta limitación.

Por otro lado, se estudiarán diferentes métodos de visualización alternativos para los historiales médicos, con el objetivo de mejorar la interpretación de los datos sin necesidad de modificar la representación actual. Estos esfuerzos buscan optimizar la comprensión y el análisis de los resultados obtenidos, para poder validarlos con los expertos del área.

4. Formación de Recursos Humanos

Durante la realización de esta investigación se espera lograr, la culminación de 2 tesis de grado dirigidas y/o codirigidas por los integrantes del proyecto.

Así también, se espera que durante el desarrollo del proyecto, los integrantes del mismo estén en proceso de elaboración de su tesis de posgrado puedan consolidar su formación en investigación, y que el trabajo realizado contribuya a su graduación.

Finalmente, es constante la búsqueda hacia la consolidación como investigadores de los miembros más recientes del grupo.

Referencias

[1] Neesha Jothi, Wahidah Husain, et al. Data mining in healthcare–a review. *Procedia computer science*, 72:306–313, 2015.

[2] Pavlos Delias, Michael Doumpos, Evangelos Grigoroudis, Panagiotis Manolitzas, and Nikolaos Matsatsinis. Supporting healthcare management decisions via robust clustering of event logs. *Knowledge-Based Systems*, 84:203–213, 2015.

[3] Ramzi A Haraty, Mohamad Dimishkieh, and Mehedi Masud. An enhanced k-means clustering algorithm for pattern discovery in healthcare data. *International Journal of distributed sensor networks*, 11(6):615740, 2015.

[4] Neerja Negi and Geetika Chawla. Clustering algorithms in healthcare. In *Intelligent*

healthcare: Applications of ai in ehealth, pages 211–224. Springer, 2021.

[5] Godwin Ogbuabor and FN Ugwoke. Clustering algorithm for a healthcare dataset using silhouette score value. *Int. J. Comput. Sci. Inf. Technol*, 10(2):27–37, 2018.

[6] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 4765–4774. Curran Associates, Inc., 2017.

[7] Scott M Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. From local explanations to global understanding with explainable ai for trees. *Nature machine intelligence*, 2(1):56–67, 2020.

[8] Sujithra Sankar and S Sathyalakshmi. A study on the explainability of thyroid cancer prediction: Shap values and association-rule based feature integration framework.

[9] Wenlong Wu, James M Keller, Marjorie Skubic, and Mihail Popescu. Explainable ai for early detection of health changes via streaming clustering. In *2022 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–6. IEEE, 2022.

Desarrollo de herramientas inteligentes para la toma de decisiones en la gestión eficiente de recursos y la seguridad informática

Carlos Bermudez[1], José Luis Hernandez[2], Mercedes Carnero[2],
Hugo Alfonso[1], Gabriela Minetti[1], Carolina Salto[1], Yanina Ditz[1]
[1]Lab. de Investigación en Sistemas Inteligentes (LISI) - Fac. de Ing. - Univ. Nac. de La Pampa (UNLPam)
e-mail: [1]minettig@ing.unlpam.edu.ar
[2]Grupo de Optimización (GOp) - Fac. de Ing. - Univ. Nac. de Río Cuarto (UNRC)

Resumen

La calidad de vida, el progreso económico y la seguridad en la sociedad actual involucran problemas de distribución de agua, de logística de distribución y de delitos informáticos. Tres problemas que pueden abordarse al combinar herramientas de la Inteligencia Artificial (metaheurísticas y *machine learning*) y cómputo paralelo, y así obtener herramientas inteligentes de soporte gerencial a la toma de decisiones.

Para una distribución eficiente del agua es importante repensar las redes de agua, tratando de minimizar costos. Por eso proponemos optimizar el diseño de redes de distribución de agua (WDND, *Water Distribution Network Design*), considerando el crecimiento poblacional.

En la logística, los costos de transporte inciden en el precio de un producto, y reducir los costos de inventario y de transporte es esencial; por eso abordamos el problema de ruteo de inventario (IRP, *Inventory Routing Problem*) para gestionar eficientemente la distribución de productos y el reabastecimiento de inventarios.

En cuanto a los delitos informáticos, amenaza creciente para la seguridad y la privacidad de los ciudadanos, se proyecta desarrollar modelos de detección de *ransomware* (*software* malicioso que bloquea dispositivos electrónicos a cambio de un rescate).

Palabras claves: Optimización, diseño redes de distribución de agua, crecimiento poblacional, gestión de inventarios, delitos informáticos

CONTEXTO

El tratamiento de problemas de interés común como son la minimización de costos de logística de distribución, la optimización del diseño de redes de distribución de agua y la reducción de delitos informáticos propenden al mejoramiento de la calidad de vida, el desarrollo económico y la seguridad de las personas. El propósito de este proyecto es aunar el conocimiento y experiencia de sus integrantes en la resolución de problemas de optimización NP-duros, usando herramientas inteligentes para brindar soluciones eficientes a las problemáticas previamente planteadas. Para esto proponemos tres líneas de trabajo donde, por separado, se estudien las tres situaciones y se desarrollen algoritmos inteligentes que den soluciones eficientes y, también, se analice el comportamiento de las propuestas algorítmicas en cuanto a su robustez, calidad de la solución y eficiencia. A continuación se describe la situación actual con respecto a los tres problemas, dando lugar a una línea de investigación para cada uno de ellos; las cuales venimos abordando conjuntamente en el LISI (Laboratorio de Investigación de Sistemas Inteligentes, UNLPam) y en el GOp (Grupo de Optimización, UNRC).

Línea de investigación 1: *Optimización del diseño de redes de distribución de agua*

En esta línea se aborda el problema de optimización del diseño de redes de distribución de agua para hacer frente al incremento en la demanda de este recurso vital. En cualquier comunidad, garantizar el acceso seguro al agua potable es de suma

importancia. Para cumplir con esta necesidad esencial, es fundamental transportar grandes cantidades de agua desde diversas fuentes, principalmente plantas de tratamiento que aseguran su calidad para el consumo humano, hacia los diferentes puntos de consumo a través de redes de distribución. Asimismo, es crucial que los diseños resultantes incluyan planes de mantenimiento adecuados. Optimizar el WDND implica considerar numerosas variables, como el tipo y diámetro de tubería, la cantidad de puntos de demanda, la presión del agua, la ubicación y capacidad de los depósitos de almacenamiento, así como los costos de los materiales de construcción, entre otros aspectos. Además, deben tenerse en cuenta diversas restricciones, como la velocidad y presión de suministro de agua, junto con restricciones mecánicas e hidráulicas.

El problema de optimizar el diseño de redes de distribución de agua ha sido clasificado como NP-duro [1], dada su inherente complejidad. Diversas estrategias, como las metaheurísticas [2]–[7], han sido empleadas para abordar versiones específicas del problema. Investigaciones más recientes se centran en variantes que consideran la variabilidad de los flujos y múltiples objetivos [8], [9]. Una aproximación destacada es el uso de metaheurísticas para optimizar el diseño considerando la variación en la demanda de agua durante las 24 horas [10].

Línea de investigación 2: *Optimización de la logística de distribución y del manejo de inventario*

El problema de ruteo de inventario se centra en optimizar la secuencia de rutas de entrega y el reabastecimiento para minimizar costos operativos, mientras se satisfacen las demandas de los clientes y se mantienen niveles adecuados de inventario. Este problema es crucial en industrias como la distribución de combustible, alimentos perecederos, comercio electrónico, entre otros. En el 2005, Guasch y Kogan [11] señalan que el desempeño logístico influye directamente en el costo de los productos y la competitividad de la industria. En tanto que, en el 2008, Guasch [12] indica que el 69% de los costos de logística están directamente relacionados con el transporte y el inventario. La gestión de inventario por sí sola representa una proporción de los activos operativos netos de aproximadamente el 37% en producción, el 62% en distribución y el 56% en venta minorista [13]. El problema de ruteo de inventarios ha sido objeto de gran atención en la comunidad científica debido a su impacto en la eficiencia y eficacia de la logística, y por su papel crucial en la sincronización y el rendimiento de la cadena de suministro [14]. El IRP implica la planificación de distribución de productos entre proveedores y clientes dispersos geográficamente durante un horizonte de planificación, con el objetivo de minimizar el costo total, que incluye tanto los costos de transporte como los de mantener el inventario. La integración del transporte y la gestión de inventarios puede resultar en ahorros significativos, con un promedio del 24% del costo total [14].

Una segunda razón por la que el IRP ganó popularidad en la comunidad científica es su capacidad para combinar el complejo problema de ruteo vehicular (VRP, *Vehicle Routing Problem*), que es NP-duro, con la gestión de inventarios y decisiones de reabastecimiento. En situaciones prácticas, la demanda y los plazos de entrega suelen ser variables, resultando en un problema estocástico complejo. Desde sus inicios, se han aplicado diferentes métodos de solución basados en inteligencia artificial para la optimización del IRP. El primer enfoque inteligente exacto fue propuesto por Archetti et al. [15] para el caso de un solo vehículo. Recientemente, la literatura se ha centrado en el caso de varios vehículos, utilizando principalmente esquemas de ramificación y corte [16]–[19]. Solo una contribución propone un algoritmo de ramificación y precio [20], con una compleja descomposición del problema.

Línea de investigación 3: *Reducción de delitos informáticos causados por ransomware*

Ransomware es un software malicioso (*malware*) que impide a los usuarios acceder a los datos de un dispositivo (o al dispositivo) y exige un pago a cambio del acceso, que generalmente se da en Bitcoin. Estos ataques aumentan en frecuencia y sofisticación, convirtiéndose en una seria amenaza para personas, empresas e instituciones.

Ante la necesidad de una detección eficaz de transacciones de *ransomware* en la red *Bitcoin*, se han utilizado métodos que identifican patrones y características asociados al *ransomware* entre un enorme número de transacciones legales. El modelo topológico de análisis de datos es un método

que hizo factible la extracción de patrones ocultos mediante un examen sistemático de las formas que adoptan los datos [21]. Pero aún es necesario superar el enfoque estático y basado en firmas. Para ello se ha propuesto el sistema dinámico de detección de *ransomware*, utilizando técnicas de *machine learning* como Random Forest, Support Vector Machine, Regresión logística, KNN y Análisis Discriminante lineal, Naive Bayes, árboles de decisión, Bagging, Ascenso de gradiente, entre otros [22]–[26].

Si bien, las técnicas mencionadas mejoran significativamente las predicciones de *ransomware*, la precisión y la rapidez de detección de direcciones *Bitcoin* aún son lentas. Por ende, se necesita crear un sistema automatizado basado en redes neuronales (otra técnica de *machine learning*) que capture la información en varias capas y detecte direcciones fraudulentas de forma generalizada, facilitando el trabajo en tiempo real.

1. DESARROLLO

En este proyecto se define como hipótesis de trabajo que el software inteligente, basado en metaheurísticas y *machine learning* para resolver las tres problemáticas ya enunciadas, sería un recurso útil para la toma eficiente de decisiones. Por ende, nuestro objetivo es aplicar el conocimiento y la experiencia de los integrantes del proyecto en la resolución de problemas de optimización NP-duros, utilizando herramientas inteligentes y, así, ofrecer soluciones eficientes que resuelvan los problemas que se describen en cada una de las líneas de investigación.

La línea 1, Optimización del diseño de redes de distribución de agua, se enfocará al diseño y evaluación de algoritmos que mejoren la eficiencia y robustez de las redes de distribución de agua. El objetivo del problema WDND es minimizar el costo total de inversión (*Total Investment Cost* - TIC) de una red de distribución de agua, que puede ser modelada con un grafo conexo conformado por un conjunto de nodos, un conjunto de tuberías, un conjunto de subredes o loops internos y un conjunto de tuberías disponibles en el mercado. En esta minimización hay que contemplar restricciones dadas por leyes físicas de conservación de masa y energía, demanda de presión mínima en cada nodo y la máxima velocidad en la cañería.

El objetivo de esta investigación es mejorar herramientas inteligentes para optimizar WDND, aumentando la eficiencia con ML y adaptando diseños a la dinámica de consumo a mediano y largo plazo. Se espera brindar un recurso útil para la toma de decisiones en el diseño de redes de agua y disminuir costos de obras públicas al seleccionar materiales adecuados para demandas presentes y futuras.

La línea 2, Optimización de la logística de distribución y del manejo de inventario, se dedicará al análisis y desarrollo de herramientas inteligentes para reducir los costos asociados a la distribución de mercadería y al manejo de inventarios. La integración de los diferentes procesos y actores que componen la Cadena de Suministro (CS) es fundamental para obtener un mejor nivel de coordinación. El control de inventarios y la gestión de la distribución son los dos procesos que los investigadores han identificado como claves para ganar o perder eficiencia y eficacia en el ámbito de la logística, con un efecto directo en la sincronización y el rendimiento general de la CS. El problema de ruteo de inventario o IRP consiste en armar un plan de distribución que cubra un horizonte de planificación determinado, donde los productos se distribuyen desde el proveedor a un conjunto de clientes geográficamente dispersos que enfrentan una demanda por período. Se dispone de una flota de vehículos para realizar las entregas y las reposiciones tienen que ser tales que cada cliente sea siempre capaz de satisfacer su demanda, en cada período del horizonte de planificación. El objetivo es determinar el plan de distribución que minimice el costo total, dado por la suma del costo de transporte (o ruta) y el costo de mantener el inventario.

El objetivo de esta línea de investigación busca desarrollar metaheurísticas para optimizar la logística de distribución de cargas y paquetería, basándose en la experiencia en VRP y sus variantes. Se evaluará el desempeño de los algoritmos con *benchmarks* científicos y su aplicación en escenarios reales. El objetivo final es crear un software inteligente para resolver el IRP y transferir este conocimiento a empresas de la región, para que mejoren la asignación de recursos y reduzcan los costos operativos.

La línea 3, Reducción de delitos informáticos causados por *ransomware*, desarrollará herramientas inteligentes basadas en ML para mitigar y prevenir delitos informáticos relacionados con la detección de *ransomware* en transacciones con *bitcoins*.

El *ransomware* es un tipo de *malware* que se propaga a sí mismo y causa serios problemas de seguridad mediante el uso de encriptación para almacenar o eliminar datos en una computadora o red. Es decir, bloquea la computadora o impide acceder a los datos, utilizando el cifrado de clave privada hasta que se pague un rescate. En la actualidad, este rescate generalmente se paga en *Bitcoin*.

El sistema de criptomonedas *Bitcoin*, basado en la tecnología *blockchain*, permite a los usuarios realizar transacciones de forma segura y pseudoanónima mediante el uso de un número arbitrario de alias (direcciones de *Bitcoin*). Los ciberdelincuentes aprovechan estas características para cometer fraude monetario inmutable y presumiblemente imposible de rastrear, especialmente mediante *ransomware*. Utilizaremos algoritmos ML relacionados con el aprendizaje supervisado y profundo para analizar cómo se llevan a cabo estas transacciones e intentaremos construir modelos que predigan si una dirección *Bitcoin* determinada se utiliza con intenciones maliciosas o no.

El objetivo de esta línea de investigación se centra en el análisis y desarrollo de algoritmos de ML, para descubrir patrones y comportamientos asociados al *ransomware* en las transacciones de *Bitcoin*. De esta manera, se espera obtener un modelo eficiente para la detección de esta clase de delito cibernético y que sea competitivo con los encontrados en la literatura.

Las soluciones algorítmicas inteligentes propuestas en las tres líneas de investigación se evaluarán en términos de robustez, calidad de la solución y eficiencia. Esta evaluación permitirá asegurar que nuestras propuestas sean tanto efectivas como aplicables en la práctica, contribuyendo significativamente a la mejora del bienestar, el crecimiento económico y la seguridad y, también, al estado del arte relacionado a estas problemáticas y a la inteligencia artificial. Además, se analizarán las posibles extensiones para la ejecución de las propuestas algorítmicas en entornos distribuidos y/o paralelos para obtener reducciones en el tiempo de ejecución (principalmente frente a casos de grandes dimensiones) o hacer frente a grandes volúmenes de datos.

2. Resultados Obtenidos

En esta sección se sintetizan logros obtenidos con relación directa o indirecta a los problemas abordados. Desde 2017, trabajamos en herramientas inteligentes basadas en metaheurísticas de trayectoria para resolver el problema WDND, mejorando iterativamente una única solución. Se han analizado distintos operadores de movimientos y parámetros para optimizar la calidad de las soluciones. Se propusieron y compararon algoritmos basados en *Tabu Search* (SOTS) y *Simulated Annealing* (HSA) para optimizar diámetros de tuberías y minimizar costos [27], [28]. SOTS emplea Oscilación Estratégica para mejorar la búsqueda, mientras que HSA incorpora un procedimiento GRASP para optimizar la red de agua. Ambos métodos fueron probados exitosamente en redes reales y *benchmarks* científicos, incluyendo una red en General Pico.

Con relación al IRP, hemos abordado variantes del VRP [29], [30] utilizando diversas metaheurísticas. Para las cuales se propusieron nuevos operadores y métodos de hibridación que resuelven casos de distintas dimensiones y complejidades. Las soluciones inteligentes propuestas resuelven de forma eficiente y eficaz los problemas antes mencionados, con aportes significativos en el estado del arte.

A partir del 2023 comenzamos el estudio de delitos cibernéticos, con particular énfasis en la detección de *ransomware* en *blockchain* [31], [32].

3. Formación de Recursos Humanos

Cada año se incorporan a los proyectos alumnos avanzados en la carrera Ingeniería en Sistemas de la UNLPam, quienes trabajan en temas relacionados a la resolución de problemas de optimización usando técnicas inteligentes, con el objeto de guiarlos en el desarrollo de sus tesinas de grado y, también, de formar futuros investigadores científicos. En tanto que, la tercera línea de investigación propende a la formación en investigación de la becaria de este proyecto; cuyo trabajo le permitirá llevar a cabo gran parte de la investigación

requerida en su tesis de maestría. Por otra parte, los docentes-investigadores involucrados realizan diversos cursos de posgrado relacionados con la temática del proyecto y con su formación docente.

REFERENCES

[1] D. Yates, A. Templeman, and T. Boffey, "The computational complexity of the problem of determining least capital cost designs for water supply networks," *Engineering Optimization*, vol. 7, no. 2, pp. 143–155, 1984.

[2] A. Alperovits and U. Shamir, "Design of optimal water distribution systems," *Water Resources Research*, vol. 13, no. 6, pp. 885–900, 1977.

[3] W. Bi, G. Dandy, and H. Maier, "Improved genetic algorithm optimization of water distribution system design by incorporating domain knowledge," *Environmental Modelling & Software*, vol. 69, pp. 370–381, 2015.

[4] M. Cunha and J. Sousa, "Hydraulic infrastructures design using simulated annealing," *Journal of Infrastructure Systems*, vol. 7, no. 1, pp. 32–39, 2001.

[5] M.-D. Lin, Y.-H. Liu, G.-F. Liu, and C.-W. Chu, "Scatter search heuristic for least-cost design of water distribution networks," *Engineering Optimization*, vol. 39, no. 7, pp. 857–876, 2007.

[6] A. Vasan and S. P. Simonovic, "Optimization of water distribution network design using differential evolution," *Journal of Water Resources Planning and Management*, vol. 136, no. 2, pp. 279–287, 2010.

[7] A. C. Zecchin, A. R. Simpson, H. R. Maier, and J. B. Nixon, "Parametric study for an ant algorithm applied to water distribution system optimization," *IEEE Transactions on Evolutionary Computation*, vol. 9, no. 2, pp. 175–191, 2005.

[8] C. Bragalli, C. D'Ambrosio, J. Lee, A. Lodi, and P. Toth, "On the optimal design of water distribution networks: a practical MINLP approach," *Optimization and Engineering*, vol. 13, no. 2, pp. 219–246, 2012.

[9] R. Farmani, G. A. Walters, and D. A. Savic, "Trade-off between total cost and reliability for anytown water distribution network," *Journal of Water Resources Planning and Management*, vol. 131, no. 3, pp. 161–171, 2005.

[10] A. De Corte and K. Sörensen, "An iterated local search algorithm for water distribution network design optimization," *Network*, vol. 67, no. 3, pp. 187–198, 2016.

[11] J. L. Guasch and J. Kogan, "Inventories and logistic costs in developing countries: levels and determinants – a red flag for competitiveness and growth," *Revista de la Competencia y la Propiedad Intelectual*, vol. 1, no. 1, 2005.

[12] J. L. Guasch, "Logistic costs in latin america and caribbean," Banco Mundial, Washington, DC, Tech. Rep., 2008.

[13] S. G. Timme and C. Williams-Timme, "The real cost of holding inventory," *Supply Chain Management Review*, vol. 7, pp. 30–37, 2003.

[14] C. Archetti and M. G. Speranza, "The inventory routing problem: The value of integration," *International Transactions in Operational Research*, vol. 23, no. 3, pp. 393–407, 2016.

[15] C. Archetti, L. Bertazzi, G. Laporte, and M. G. Speranza, "A branch-and-cut algorithm for a vendor-managed inventory-routing problem," *Transportation Science*, vol. 41, no. 3, pp. 382–391, 2007.

[16] Y. Adulyasak, C. Archetti, and M. G. Speranza, "Optimization methods for the irp: Branch-and-cut schemes," *Journal of Optimization*, 2014.

[17] P. Avella, M. Boccia, A. Sforza, and C. Sterle, "Multi-vehicle inventory routing: An overview," *Journal of Industrial Engineering*, 2018.

[18] L. C. Coelho, J.-F. Cordeau, and G. Laporte, "Branch-and-cut methods for irp," *Transportation Science*, 2014.

[19] K. Manousakis, P. Repoussis, and C. Tarantilis, "Recent advances in irp solutions," *Journal of Operations Management*, 2021.

[20] G. Desaulniers, F. Errico, and W. Rei, "Branch-and-price algorithm for irp," *Mathematical Programming*, 2016.

[21] C. G. Akcora, Y. Li, Y. R. Gel, and M. Kantarcioglu, "Bitcoin heist: Topological data analysis for ransomware detection on the bitcoin blockchain," *arXiv preprint*, 2019.

[22] F. Karbalaie, A. Sami, and M. Ahmadi, "Semantic malware detection by deploying graph mining," *International Journal of Computer Science Issues (IJCSI)*, vol. 9, no. 1, p. 373, 2012.

[23] A. Kharraz, S. Arshad, C. Mulliner, W. K. Robertson, and E. Kirda, "Unveil: a large-scale, automated approach to detecting ransomware," in *Usenix security symposium*, vol. 25, 2016.

[24] Z.-G. Chen, H.-S. Kang, S.-N. Yin, and S.-R. Kim, "Automatic ransomware detection and analysis based on dynamic api calls flow graph," in *Proceedings of the international conference on research in adaptive and convergent systems*, 2017, pp. 196–201.

[25] H. Sun Yin and R. Vatrapu, "A first estimation of the proportion of cybercriminal entities in the bitcoin ecosystem using supervised machine learning," in *2017 IEEE International Conference on Big Data (Big Data)*, 2017, pp. 3690–3699.

[26] P. Nerurkar, "Illegal activity detection on bitcoin transaction using deep learning," *Soft Computing*, pp. 1–18, 2023.

[27] C. Salto, G. Minetti, H. Alfonso, E. Zurita, J. Hernández, and M. Carnero, "Optimal design of water distribution network using two different metaheuristics," *CHEMICAL ENGINEERING TRANSACTIONS*, vol. 100, 2023.

[28] C. Bermudez, H. Alfonso, G. Minetti, and C. Salto, "A parallel optimization solver for the multi-period wdnd problem," in *Hybrid Artificial Intelligent Systems. HAIS 2021*, ser. Lecture Notes in Computer Science, H. Sanjurjo González, I. Pastor López, P. García Bringas, H. Quintián, and E. Corchado, Eds. Springer, Cham, 2021, vol. 12886.

[29] C. Bermúdez, H. Alfonso, G. Minetti, and C. Salto, "Adapting pals to solve vrp with time windows," in *Congreso Argentino de Ciencias de la Computación CACIC 2023*. Argentina: Universidad Nacional de Luján, 2023. [Online]. Available: https://cacic2023.unlu.edu.ar/congreso/papersAceptados.html

[30] P. Graglia, N. Stark, C. Salto, and H. Alfonso, "A comparison of recombination operators for capacitated vehicle routing problem," in *Libro de Actas del XIV Congreso Argentino de Ciencias de la Computación (CACIC 2008)*, Chilecito, La Rioja, Argentina, oct 2008, pp. 460–469.

[31] Y. Ditz and G. Minetti, "Detección de ransomware en blockchains," in *Actas del WorkShop de Investigadores en Ciencias de la Computación WICC 2023*, Argentina: Red UNCI y Universidad Nacional del Noroeste de la Prov. de Bs. As., 2023.

[32] ——, "Detección de ransomware en red bitcoin," in *L Coloquio Argentino de Estadística, VIII Jornadas de Educación Estadística "Martha Aliaga"*. Universidad Nacional de Cuyo, 2023. [Online]. Available: https://sites.google.com/view/lcaeviiijee/actividades/programa

DESARROLLO DE TÉCNICAS DE INTELIGENCIA COMPUTACIONAL Y ARTIFICIAL APLICADAS A PROBLEMAS GENÓMICA ESTRUCTURAL Y FUNCIONAL

Pablo Javier Vidal[a,b], Jessica Andrea Carballido[c], Ana Carolina Olivera [a,b], Matías Gabriel Rojas[a] e Ignacio Bosch[d]

[a]*Instituto Interdisciplicario de Ciencias Básicas Consejo Nacional de Investigaciones Científicas y Técnicas. Universidad Nacional de Cuyo (ICB CONICET UNCuyo)*

[b]*Facultad de Ingeniería - Universidad Nacional de Cuyo (FING-UNCuyo)*

[c]*Instituto de Ciencias e Ingeniería de la Computación, Consejo Nacional de Investigaciones Científicas y Técnicas, Universidad Nacional del Sur (ICIC-CONICET-UNS)*

[d]*Facultad de Ciencias Exactas y Naturales de la Universidad Nacional de Cuyo (FCEN-UNCuyo)*

contacto: pjvidal@conicet.gov.ar,pablo.vidal@ingenieria.uncuyo.edu.ar

CONTEXTO

Esta presentación corresponde a las tareas de investigación que se llevan a cabo en el Grupo de Ciencias de la Computación y de los Datos en el Instituto Interdisciplinario de Ciencias Básicas (ICB CONICET UNCuyo) en conjunto con la Facultad de Ingeniería de la UNCuyo y la Universidad Nacional del Sur en el marco del proyecto 06/B052-T1 (2022-2025), un Proyecto de investigación bianual para Investigadoras/es Asistentes y Adjuntas/os de reciente ingreso al CONICET (PIBAA 2022-2023) y una Beca Doctoral del CONICET.

RESUMEN

Esta línea de investigación se centra en el diseño y desarrollo de técnicas de Inteligencia Computacional y Artificial en combinación con otros métodos con el propósito de contribuir al área de Ciencias de la Computación aplicando el conocimiento desarrollado a problemas de bioinformática, en particular aquellos de las áreas de genómica estructural y funcional.

Palabras clave: Bioinformática, Genómica Estructural, Genómica Funcional, Inteligencia Computacional, Inteligencia Artificial.

1 INTRODUCCIÓN

El conocimiento de las secuencias del genoma humano y de otras especies permite que se realicen investigaciones del genoma como un todo. Para analizar el conjunto de genes la bioinformática trabaja con diversas técnicas computacionales que están a disposición de los científicos debido a que pueden procesar e interpretar una gran cantidad de datos en n-dimensiones y reconstruir y separar estos conjuntos para tareas de clasificación o generar regresiones numéricas para tareas de predicción. Esto permite determinar automáticamente si un caso de estudio presenta una patología especifica o bien, si ciertas características del DNA (Deoxyribonucleic Acid) pueden permitir una clasificación de nuevos casos en forma rápida y eficiente, entre otras posibilidades.

Entre los problemas abordados por el grupo de investigación se encuentran la selección de genes en conjuntos de datos de expresiones génicas.Asimismo, se desarrollaron nuevos operadores para técnias de inteligencia computacional que son combinadas con inteligencia artificial para poder clasificar patologías en conjuntos de datos de enfermedades.

2 LINEAS DE INVESTIGACIÓN y DESARROLLO

- Diseño de novedosas técnicas de Inteligencia Computacional y artificial hibridizando y combinándolas con otras técnicas.

- Aplicación de Inteligencia Computacional y Artificial a problemas de Bioinformática y salud.
- Ajuste de hiperparámetros de redes neuronales artificiales y aplicación a problemas de bioinformática.

3 RESULTADOS OBTENIDOS/ESPERADOS

3.1 Diseño de nuevos operadores e hibridación de metaheurísticas

La tecnología en medicina ha mostrado un avance significativo debido a que la inteligencia artificial se ha convertido en un marco para realizar diagnósticos médicos precisos. Los modelos como los perceptrones multicapa (MLP, en inglés MultilayerPerceptron) pueden detectar patrones implícitos en los datos, lo que permite identificar las condiciones de los pacientes que no se pueden ver fácilmente. Un MLP consiste en neuronas sesgadas dispuestas en capas, conectadas por conexiones ponderadas. Su efectividad depende de encontrar los pesos y sesgos óptimos que reduzcan el error de clasificación, lo que generalmente se hace utilizando el algoritmo Back Propagation (BP). Pero BP tiene varias desventajas que podrían provocar que el MLP no aprenda. Las metaheurísticas son alternativas a BP que alcanzan soluciones de alta calidad sin utilizar muchos recursos computacionales. Se realizaron varias aproximaciones a la resolución del problema mencionado.

Se diseñaron diferentes metaheurísticas y, en particular, se comenzó con una propuesta para el algoritmo genético celular. Para este se construyó un nuevo operador de cruzamiento inspirado en el Oscilador Armónico. Este operador permite en las iteraciones iniciales del algoritmo genético celular realizar una búsqueda acelerada para luego ir disminuyendo a una búsqueda más intensiva del espacio subyacente a las soluciones de las iteraciones finales. Como escenarios de análisis se utilizaron 5 datasets (Cáncer de Mama, Diabetes, Cáncer de Hígado, Parkinsons, Columna Vertebral) y 8 algoritmos para comparación. Del análisis surge que nuestro operador mejora notablemente el algoritmo canónico y sobrepasa a casi todos los algoritmos para todos los conjuntos de datos [ROV22].

Asimismo, también se realizó una aproximación de varios objetivos con un algoritmo multiobjetivo basado en el Enjambre de Partículas para optimizar los pesos y sesgos del Perceptrón Multicapa para clasificación de datos médicos [RCOV22a]. Siguiendo con esta línea de investigación se ha propuesto una variación híbrida del Algoritmo Micro-Genético (mGA, micro Genetic Algorithm) como un algoritmo híbrido con un método de búsqueda local específicamente diseñado para el problema de selección de genes determinantes en conjuntos de datos de expresiones génicas en microarrays [ROCV23].

En lo que respecta a la mejora de hiperparámetros se construyó una variante genética del algoritmo de optimización de hormigas león diseñando un operador de cruzamiento específico que tiene por objetivo controlar las contribuciones de las hormigas león en la actualización de las posiciones de las hormigas comunes y evitar la pérdida de información (dos de los principales problemas del algoritmo de optimización de hormigas león) [ROV24].

4 FORMACIÓN DE RECURSOS HUMANOS

El grupo de trabajo está formado por: Un Profesor Titular con Dedicación Simple de la Facultad de Ingeniería de la UNCuyo (Investigador Adjunto del CONICET), una Profesora Adjunta con Dedicación Exclusiva de la Universidad Nacional del Sur (Investigadora Independiente del CONICET), una Profesora Titular con Dedicación Semiexclusiva de la Facultad de Ingeniería

de la UNCuyo (Investigadora Independiente del CONICET), un Becario doctoral del CONICET y un doctorando de la Facultad de Ciencias Exactas y Naturales de la Universidad Nacional de Cuyo.

El Ing. Rojas es dirigido por el Dr. Vidal y la Dra. Carballido en su doctorado en la Facultad de Ciencias Exactas y Naturales de la UNCuyo en el tema "*Desarrollo de Metaheurísticas Aplicadas a Genómica Funcional y Estructural*".

El Ing. Ignacio Bosch es dirigido por los Dres. Vidal y Olivera en su doctorado en la Facultad de Ciencias Exactas y Naturales de la UNCuyo en el tema "*Diseño y Desarrollo de Técnicas de Inteligencia Computacional aplicadas a la detección de biomarcadores para Prognosis en Huntington".*

5 BIBLIOGRAFIA

[RCOV22] Matías Gabriel Rojas, Jessica Andrea Carballido, Ana Carolina Olivera, and Pablo Javier Vidal. A modified multiswarm optimization for weights and biases of a multi-layer perceptron for medical data classification. In 2022 IEEE Biennial Congress of Argentina (ARGENCON), pages 1–8, 2022. doi:10.1109/ARGENCON55245.2022.9940066.

[ROV22] Matías Gabriel Rojas, Ana Carolina Olivera, and Pablo Javier Vidal. Optimising multilayer perceptron weights and biases through a cellular genetic algorithm for medical data classification. Array, 14:100173, 2022. URL: https://doi.org/10.1016/j.array.2022.100173.

[ROCV23] Matías Gabriel Rojas, Ana Carolina Olivera, Jessica Andrea Carballido, and Pablo Javier Vidal. Memetic microgenetic algorithms for cancer data classification. Intelligent Systems with Applications, 17:200173, 2023. URL: https: //doi.org/10.1016/j.iswa.2022.200173.

[ROV24] Matías Gabriel Rojas, Ana Carolina Olivera, and Pablo Javier Vidal. A genetic operators-based Ant Lion Optimiser for training a medical multi-layer perceptron 2024 Applied Soft Computing, Vol. 151 p. 111192

[ROV25] Matías Gabriel Rojas, Ana Carolina Olivera, and Pablo Javier Vidal. A memetic genetic particle swarm optimization for druglike molecule discovery. IEEE LATIN AMERICA TRANSACTIONS, 23 (3):216–222, March 2025 https://doi.org/10.1109/tla.2025.10879174

EcoConversa: Evolución hacia el aprendizaje por refuerzo para la promoción de los ODS

F. Brandan, A. Villagra, D. Pandolfi
{fbrandan, avillagra, dpandolfi}@uaco.unpa.edu.ar

[1] Laboratorio de Tecnologías Emergentes (LabTEm)
Instituto de Tecnología Aplicada (ITA) - Unidad Académica Caleta Olivia
Universidad Nacional de la Patagonia Austral

RESUMEN

La Inteligencia Artificial (IA) ha demostrado su potencial en la resolución de problemas globales, incluyendo los Objetivos de Desarrollo Sostenible (ODS) que plantea la ONU. El proyecto EcoConversa busca fortalecer el compromiso con estos objetivos mediante un agente conversacional basado en técnicas de Procesamiento de Lenguaje Natural (NLP) y aprendizaje automático.

En la iteración actual del proyecto se ha explorado el Aprendizaje por Refuerzo (RL) como una alternativa al enfoque supervisado tradicional, con el objetivo de mejorar la autonomía y adaptabilidad del chatbot. Se inició la implementación de Q-Learning en un entorno controlado (juego de Pong) para comprender las metodologías de entrenamiento en este paradigma y posteriormente incorporarlas en el agente conversacional. Este trabajo presenta una comparación entre RL y otros enfoques de aprendizaje, así como los avances y desafíos en la aplicación de estas técnicas en EcoConversa.

Palabras clave: Chatbot; Aprendizaje por Refuerzo; Procesamiento del Lenguaje Natural; Objetivos de Desarrollo Sostenible, Inteligencia Artificial.

CONTEXTO

La línea de trabajo del proyecto EcoConversa se desarrolla en el Laboratorio de Tecnologías Emergentes (LabTEm) del Instituto de Tecnología Aplicada (ITA) de la Universidad Nacional de la Patagonia Austral, Unidad Académica Caleta Olivia, en el marco del Proyecto de Investigación "Enfoques inteligentes para ciudades del futuro: Empleo, Movilidad y Ambiente Sostenible".

1. INTRODUCCIÓN

Los chatbots han probado su versatilidad en múltiples áreas, desde el servicio al cliente hasta la educación y la concientización social. Dentro del marco del proyecto EcoConversa el objetivo es desarrollar un agente de conversación que ayude a fomentar los Objetivos de Desarrollo Sostenible, poniendo especial atención a los ODS 11 (Ciudades y Comunidad Sostenibles) y 17 (Alianzas para alcanzar los objetivos). La meta principal del agente es facilitar el acceso a información pertinente sobre los ODS, incentivando la concientización y promoviendo el compromiso ciudadano con estos objetivos a nivel mundial. El agente puede contribuir a la propagación de saberes sobre sostenibilidad, atender preguntas sobre buenas prácticas ambientales y vincular a las personas con los recursos necesarios para actuar.

Un chatbot es un sistema diseñado para replicar diálogos humanos a través de Inteligencia Artificial. Por lo general se compone de un módulo de comprensión de lenguaje natural, un motor para la toma de decisiones y un generador de respuestas. Mediante el reconocimiento de lenguaje natural, el chatbot comprende la intención de las peticiones del usuario y elige la respuesta más adecuada, mientras que las Redes Neuronales Artificiales (ANN) pueden ser utilizadas para potenciar su habilidad de aprendizaje y generalización.

Inicialmente, EcoConversa se desarrolló utilizando un enfoque supervisado basado en Rasa, un *framework*/marco de trabajo que permite la creación de agentes conversacionales de forma versátil y con habilidad de reconocimiento de intenciones de forma nativa.

Con el objetivo de enseñarle el lenguaje español se le brindó un corpus lingüístico en español (AnCora-ES3), una colección de artículos periodísticos diversos, y una serie de mensajes que podría recibir de los usuarios. Rasa utiliza clasificadores de NLP (*Natural Language Processing*) que examinan el texto del usuario y lo vinculan con la intención más probable, haciendo uso de *embeddings* de palabras. Por ejemplo, si un usuario cuestiona "¿Cómo puedo disminuir mi huella de carbono?", el sistema puede reconocer la intención "consulta_sostenibilidad" y escoger una respuesta apropiada.

No obstante, este método supervisado tiene restricciones respecto a la capacidad del sistema para adaptarse a nuevos contextos y estilos de diálogo. En la etapa actual del proyecto, se ha iniciado el estudio del Aprendizaje por Refuerzo como una táctica para potenciar la habilidad del chatbot para aprender y tomar decisiones. Un caso de uso de estas técnicas en el presente es ChatGPT, que fusiona modelos de aprendizaje profundo con RL mediante técnicas como RLHF (*Reinforcement Learning from Human Feedback*, o Aprendizaje por Refuerzo a partir de Retroalimentación Humana) para mejorar la calidad de sus respuestas. Otros sistemas, como Copilot de GitHub, también utilizan métodos sofisticados de NLP y RL para optimizar su rendimiento en labores de ayuda basadas en Inteligencia Artificial.

RL facilita que un agente adquiera conocimientos a través de la interacción con su ambiente, recibiendo recompensas por acciones que lo acercarán a su meta. En contraposición al aprendizaje supervisado, que demanda un conjunto de datos etiquetado, el RL posibilita que los agentes optimicen su rendimiento a través de prueba y error, lo cual es particularmente beneficioso para ambientes dinámicos y de conversación natural.

En la actualidad, EcoConversa está en una etapa de transición hacia la utilización de RL, utilizando un modelo experimental basado en Q-Learning. Se optó por utilizar esta metodología como primera etapa de adaptación, gracias a su sencillez y efectividad en ambientes discretos, facilitando un entendimiento progresivo de los principios fundamentales del Aprendizaje por Refuerzo. Para tal fin, se creó un ambiente simplificado basado en el juego de Pong, completamente desarrollado en Python desde el inicio. Esta implementación posibilitó la creación de un marco regulado en el que el agente pudiera adquirir estrategias de juego a través de la experimentación, mejorando su política de acción basada en la acumulación de recompensas. La experiencia obtenida en este contexto se utilizó como fundamento para la futura incorporación del RL en el chatbot, simplificando el paso hacia modelos de toma de decisiones más sofisticados en contextos de conversación.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El trabajo actual en EcoConversa se enfoca en la transición de una metodología supervisada a una basado en RL, con la finalidad de potenciar la capacidad de adaptación y autonomía del chatbot en sus interacciones. Esta transformación conlleva una reorganización del modelo de aprendizaje, posibilitado que el agente optimice sus respuestas de manera dinámica y en función de la retroalimentación recibida durante su uso. Para entender de manera más profunda este proceso, es esencial examinar los tres paradigmas principales de aprendizaje por IA:

- **Aprendizaje supervisado:** Requiere un conjunto de datos etiquetado y utiliza estos ejemplos para entrenar modelos capaces de hacer predicciones precisas en datos nuevos. Se emplea frecuentemente en el entrenamiento de modelos de clasificación y NLP.
- **Aprendizaje no supervisado:** No depende de etiquetas, sino que busca patrones y estructuras ocultas en los datos. Se utiliza en aplicaciones como la agrupación de textos o la detección de anomalías.
- **Aprendizaje por refuerzo:** Un agente interactúa con su ambiente, realizando acciones y obteniendo recompensas en fun-

ción de su desempeño. A través de la optimización de una política, el agente aprende a maximizar su recompensa a largo plazo.

El RL se basa en el Proceso de Decisión de Markov (MDP), que define formalmente el ambiente de aprendizaje mediante los siguientes elementos:

- **Estados (S):** Representan la situación actual del agente en el entorno.
- **Acciones (A):** Conjunto de movimientos/operaciones posibles que el agente puede realizar.
- **Recompensas (R):** Valores que indican qué tan beneficiosa fue una acción tomada en un estado específico.
- **Política (π):** Estrategia que guía la selección de acciones basadas en el estado actual.

En el caso de EcoConversa, el chatbot puede ser modelado como un MDP en el que:

- Los **estados** corresponden a las interacciones previas y el contexto de la conversación actual.
- Las **acciones** contemplan la selección de respuestas y recomendaciones de contenido.
- Las **recompensas** se definen en función de indicadores como la relevancia de la respuesta y la satisfacción del usuario.

Para avanzar en la implementación de RL en EcoConversa, se han realizado las siguientes etapas:

- **Exploración de RL en entornos simples:** Aplicación de un algoritmo de Q-Learning en un juego de Pong, permitiendo al agente aprender a jugar a partir de la optimización de su política de acciones.
- **Adaptación de RL al contexto conversacional:** Análisis de arquitecturas de RL aplicadas a chatbots, tales como Deep Q-Networks (DQN) y Proximal Policy Optimization (PPO), con el objetivo de evaluar su aplicabilidad en la mejora de la interacción con los usuarios.
- **Definición del espacio de estados y acciones:** Identificación de las posibles entradas y salidas del chatbot, además de una función de recompensa que promueva respuestas informativas, contextualizadas y satisfactorias para el usuario.

Los retos más significativos abarcan la necesidad de definir recompensas adecuadas en contextos de conversación y la escalabilidad del modelo para interactuar con usuarios de diversos grados de saber. Se plantea la hipótesis que la aplicación de RL en EcoConversa permitirá mejorar la adaptabilidad del chatbot, mejorando la selección de respuestas en función de la retroalimentación del usuario y promoviendo interacciones más fluidas y con mayor relevancia en el contexto.

En particular, este trabajo está enfocado en la creación e implementación de un chatbot funcional a través de la red social Instagram y página web del laboratorio donde se lleva a cabo el proyecto (LabTEm), completamente orientado a la respuesta informativa sobre los ODS, la propuesta de la ONU y sugerir posibles acciones para aportar a la agenda del desarrollo sostenible.

3. RESULTADOS OBTENIDOS/ESPERADOS

El resultado esperado del proyecto es la puesta en marcha de la inteligencia en un medio en línea para que cualquier usuario interesado en la sostenibilidad y los Objetivos de Desarrollo Sostenible pueda interactuar con ella de manera libre. Para alcanzar este objetivo, se espera que el chatbot no solo comprenda las consultas y produzca respuestas apropiadas, sino que también mejore su desempeño a través del Aprendizaje por Refuerzo. Este enfoque facilitará la mejora de la adaptabilidad del agente en contextos dinámicos, maximizando su capacidad de aprendizaje en situaciones donde el conocimiento del modelo de lenguaje es limitado.

Los progresos logrados hasta el momento incluyen:

- Desarrollo exitoso de Q-Learning en un entorno simple, comprendiendo la dinámica del entrenamiento mediante RL.
- Elaboración de la arquitectura inicial para incorporar RL en EcoConversa, con la finalidad de potenciar la capacidad de adaptación del chatbot a diferentes interacciones.
- Desarrollo de estrategias para mediar la calidad de las respuestas generadas por el chatbot bajo la nueva metodología.

El proyecto sigue en un estado de desarrollo dinámico y progresivo, sometido a iteraciones constantes y mejoras iterativas.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo consta de tres investigadores de diferentes niveles de formación, una Doctora en Ciencias de la Computación, un Magister en Ciencias de la Computación, y un estudiante avanzado de la carrera Ingeniería en Sistemas.

Esta línea de investigación ofrecerá un contexto favorable para el inicio y/o culminación de estudios de grado y de posgrado de los integrantes del Laboratorio.

BIBLIOGRAFÍA

Brandan, F., Villagra, A., & Pandolfi, D. (2023). Sistema inteligente conversacional aplicado a la gobernanza local: Automatización de asistencia pública. *Informes Científicos - Técnicos UNPA*, *15*(3), 49–69. https://doi.org/10.22305/ict-unpa.v15.n3.983

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

Jason, D. W., & Geoffrey, Z. (2016). End-to-end LSTM-based dialog control optimized with supervised and reinforcement learning. En *arXiv [cs.CL]*. http://arxiv.org/abs/1606.01269

Jiao, A. (2020). An Intelligent chatbot system based on entity extraction using RASA NLU and neural network. *Journal of physics. Conference series*, *1487*(1), 012014. https://doi.org/10.1088/1742-6596/1487/1/012014

Jurafsky, D. (s/f). *Speech and Language Processing*. Stanford.edu. Recuperado de https://web.stanford.edu/~jurafsky/slp3/

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2021). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. En *arXiv [cs.CL]*. http://arxiv.org/abs/2107.13586

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529–533. https://doi.org/10.1038/nature14236

Pearl, J. (1984). *Heuristics: Intelligent search strategies for computer problem solving*. Addison Wesley Longman Publishing.

Rasa: Developer documentation portal. (s/f). Rasa. Recuperado de https://rasa.com/docs

Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4a ed.). Pearson.

Sid, M., Vineet, G., Frederick, L., & Mukund, S. (2023). Using foundation models to detect policy violations with minimal supervision. En *arXiv [cs.CL]*. http://arxiv.org/abs/2306.06234

Sutton, R. S., & Barto, A. G. (2012). *An Reinforcement Learning: Introduction*. Mit Press.

Take action for the sustainable development goals. (2018, mayo 17). United Nations Sustainable Development; United Nations: Sustainable Development Goals. https://www.un.org/sustainabledevelopment/sustainable-development-goals.

ENFOQUES INTELIGENTES PARA MOVILIDAD Y MEDIOAMBIENTE: TRANSPORTE MULTIMODAL Y GESTIÓN DE RESIDUOS COSTEROS

Villagra A[1]., Valdez J[1]., Luque G[4]., Pandolfi D[1]., Brandan F[1]., Errecalde M[1,2]. Villagra A.M[1], Gonzalez C[3].

[1]Laboratorio de Tecnologias Emergentes (LabTEm)
Instituto de Tecnologia Aplicada (ITA) - Unidad Académica Caleta Olivia
Universidad Nacional de la Patagonia Austral

[2]Laboratorio de Investigación y Desarrollo en Inteligencia Computacional (LIDIC)
Departamento de Informática - Universidad Nacional de San Luis

[3]Unidad Académica Río Gallegos, Universidad Nacional de la Patagonia Austral)

[4]Grupo de Investigación NEO, Universidad de Málaga (España)

{avillagra, jcvaldez}@uaco.unpa.edu.ar, gluque@uma.es, {dpandolfi, fbrandan}@uaco.unpa.edu.ar, merreca@unsl.edu.ar, cgonzalez@uarg.unpa.edu.ar

RESUMEN

Este artículo presenta dos líneas de investigación clave que abordan desafíos urbanos contemporáneos en movilidad y sostenibilidad ambiental mediante el uso de enfoques inteligentes. En el ámbito de la movilidad, se propone un sistema de recomendación de transporte multimodal que busca optimizar la elección de modos de transporte para los usuarios, considerando tanto opciones unimodales como multimodales. En el ámbito ambiental, se desarrolla un sistema de identificación de residuos costeros mediante técnicas de reconstrucción de imágenes con aprendizaje profundo, que permite la detección y clasificación eficiente de desechos en entornos costeros. Ambas líneas de investigación buscan mejorar la calidad de vida en las ciudades, reducir la congestión vehicular y promover la sostenibilidad ambiental mediante el uso de tecnologías avanzadas de inteligencia artificial.

Palabras clave: Movilidad multimodal, Identificación de residuos, Reconstrucción de imágenes, Aprendizaje profundo, Ciudades inteligentes.

CONTEXTO

La rápida urbanización y el aumento de la población en las ciudades han generado desafíos significativos en términos de movilidad y gestión de residuos. En este contexto, las ciudades inteligentes surgen como una solución para abordar estos problemas mediante el uso de tecnologías avanzadas. Este artículo se enmarca en el proyecto de investigación " Enfoques inteligentes para ciudades del futuro: Empleo, Movilidad y Ambiente Sostenible", desarrollado en el Laboratorio de Tecnologías Emergentes (LabTEm) del Instituto de Tecnología Aplicada (ITA) de la Universidad Nacional de la Patagonia Austral, en colaboración con el LIDIC de la UNSL y el Grupo NEO de la UMA (España).

1. INTRODUCCIÓN

Las ciudades inteligentes representan un paradigma emergente en la planificación urbana, cuyo objetivo principal es mejorar la calidad de vida de sus habitantes mediante la implementación de tecnologías innovadoras que optimicen

los servicios urbanos y promuevan la sostenibilidad (Toli & Murtagh, 2020). En este contexto, la movilidad y la gestión de residuos se han convertido en áreas críticas debido a su impacto directo en la eficiencia urbana y la preservación del medio ambiente.

En el ámbito de la movilidad, el aumento de la congestión vehicular y la necesidad de reducir las emisiones de carbono han impulsado la adopción de enfoques multimodales que combinan diversos medios de transporte, como autobuses, metro, bicicletas y taxis (Xu et al., 2023). Estos sistemas no solo optimizan los desplazamientos de los usuarios, sino que también ayudan a reducir la congestión y mejorar la eficiencia del transporte urbano. Sin embargo, elegir el modo de transporte más adecuado puede ser complejo, especialmente en ciudades con múltiples opciones y condiciones de tráfico variables. Para abordar este desafío, se han desarrollado sistemas de recomendación basados en inteligencia artificial que analizan grandes volúmenes de datos de movilidad y ofrecen sugerencias personalizadas (Xu et al., 2023). Estos sistemas consideran factores como el tiempo, el costo, las preferencias individuales y las condiciones del tráfico en tiempo real, facilitando una movilidad más eficiente y sostenible.

Por otro lado, la gestión de residuos, especialmente en áreas costeras, es un desafío global que requiere soluciones innovadoras para preservar los ecosistemas marinos y reducir la contaminación. La identificación y clasificación de residuos costeros es un paso crucial para implementar políticas efectivas de gestión de residuos (Borrelle et al., 2020; Lambert & Wagner, 2020). Sin embargo, la detección manual de residuos es un proceso lento y costoso, lo que ha llevado a la exploración de técnicas automatizadas basadas en visión artificial y aprendizaje profundo.

Este artículo propone dos líneas de investigación para abordar desafíos urbanos de manera integral. Primero, un sistema de recomendación de transporte multimodal que emplea inteligencia artificial para analizar datos de movilidad y sugerir el mejor medio de transporte para cada usuario. Segundo, un sistema de identificación de residuos costeros basado en técnicas de reconstrucción de imágenes con aprendizaje profundo, que detecta y clasifica desechos de manera eficiente en entornos costeros (Cheng et al., 2020; Li et al., 2022), demostrando alta precisión y facilitando políticas efectivas de gestión de residuos.

Ambas investigaciones buscan mejorar la calidad de vida urbana, reducir la congestión vehicular y promover la sostenibilidad ambiental mediante el uso de inteligencia artificial. Estas soluciones tienen el potencial de transformar la gestión del transporte y los residuos, contribuyendo a un futuro más sostenible y eficiente.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

En esta sección se describen brevemente las líneas de investigación que se llevan a cabo en el proyecto.

2.1. Recomendación de transporte multimodal

El campo de la recomendación de transporte ha evolucionado desde sistemas unimodales que utilizan algoritmos de búsqueda de rutas más cortas (Fu & Rilett, 2006), hasta enfoques multimodales que consideran múltiples combinaciones de transporte (Liu et al., 2019), (Liu et al., 2023). Inicialmente, la planificación de rutas se basaba en criterios como tiempo, costo y distancia, pero estos métodos no captaban la diversidad de preferencias de los usuarios (Rogers & Langley, 1998). Para abordar esta limitación, se han desarrollado modelos que integran personalización en la recomendación de rutas (Dai et al., 2015), junto con enfoques basados en aprendizaje automático, que optimiza el transporte público (Liu et al., 2020). Sin embargo, estos modelos aún presentan desafíos al ignorar patrones de movilidad de los usuarios (Wang et al., 2019).

La recomendación de transporte multimodal ha ganado atención en la academia (Campigotto et al., 2016) y la industria (Liu et al., 2019), al permitir combinaciones como autobús-metro o taxi-bicicleta.

Las redes neuronales de grafos (GNNs) han surgido como una solución efectiva para modelar datos de transporte (Scarselli et al., 2008), permitiendo representar relaciones complejas entre usuarios, rutas y modos de transporte.

El desarrollo de modelos capaces de manejar datos heterogéneos sigue siendo un desafío clave en la recomendación de transporte multimodal. La integración de redes neuronales con mecanismos de atención jerárquica y representación de relaciones heterogéneas se presenta como una dirección prometedora para mejorar la precisión y personalización de las recomendaciones de transporte (Xu et al., 2023).

En este contexto, se propone un sistema de recomendación de transporte multimodal que utiliza técnicas avanzadas de inteligencia artificial para analizar grandes volúmenes de datos de movilidad y recomendar el modo de transporte más adecuado para cada usuario.

En una primera fase, se trabajará con modelos clásicos de optimización y aprendizaje automático para establecer una base sólida y comprender las dinámicas básicas de la movilidad urbana. Estos modelos incluirán algoritmos para la planificación de rutas y métodos de regresión para predecir tiempos de viaje y costos (Li & Schonfeld, 2015). Estos enfoques permitirán una implementación inicial rápida y una validación preliminar del sistema.

Posteriormente, se avanzará hacia modelos más sofisticados inspirados en el enfoque THAN (Xu et al., 2023), que utiliza redes de atención heterogéneas para capturar las relaciones complejas entre usuarios, pares origen y destino (OD) y modos de transporte. Este enfoque se basa en la construcción de un grafo heterogéneo que representa las interacciones entre diferentes tipos de nodos y aristas. A través de un mecanismo de atención jerárquica, el sistema es capaz de distinguir la importancia de diferentes nodos y aristas, lo que permite una recomendación más precisa y personalizada.

2.2. Identificación de residuos costeros mediante reconstrucción de imágenes con aprendizaje profundo

Este trabajo busca desarrollar un sistema basado en inteligencia artificial que permita mejorar la detección y clasificación de desechos en entornos costeros mediante técnicas avanzadas de compresión y reconstrucción de imágenes. El problema central radica en la contaminación ambiental generada por desechos plásticos, orgánicos y peligrosos, los cuales afectan negativamente los ecosistemas marinos y la biodiversidad (Borrelle et al., 2020). La gestión eficiente de estos residuos es clave para la sostenibilidad ambiental, y la identificación precisa de los mismos es fundamental para la aplicación de políticas efectivas (United Nations Environment Programme - UNEP, 2023).

El desarrollo de sistemas de visión artificial para la clasificación de residuos ha cobrado gran relevancia en los últimos años gracias al avance del aprendizaje profundo. Modelos como YOLO (You Only Look Once) han permitido la detección de objetos en tiempo real con alta precisión, abriendo la posibilidad de automatizar el proceso de identificación de residuos (Bochkovskiy et al., 2020; Redmon et al., 2016; Redmon & Farhadi, 2018; Wang et al., 2023). Sin embargo, la captura y transmisión de imágenes en entornos remotos presenta desafíos técnicos relacionados con el ancho de banda y la capacidad de almacenamiento. La compresión de imágenes basada en técnicas de aprendizaje profundo ofrece una solución potencialmente más eficiente y adaptativa en comparación con los métodos tradicionales (Cheng et al., 2020; Li et al., 2022; Mishra et al., 2022; Yasin & Abdulazeez, 2021).

El objetivo principal de esta línea es desarrollar un sistema que combine modelos de compresión y reconstrucción de imágenes con aprendizaje profundo, mejorando la calidad visual de las imágenes transmitidas y optimizando

el proceso de detección de residuos costeros. Para ello, se propone el uso de autoencoders y redes neuronales convolucionales para reducir la carga de datos transmitidos sin comprometer la fidelidad de la información visual. Además, se empleará un modelo de detección basado en YOLO para identificar distintos tipos de residuos en las imágenes reconstruidas. Este enfoque permitirá generar un sistema robusto y eficiente para la gestión y monitoreo de residuos en entornos costeros, integrando tecnologías de *edge computing* y drones para la captura de imágenes.

La validación del sistema se llevará a cabo en diversos entornos costeros, considerando variaciones en iluminación, oclusión y tipos de residuos presentes.

3. RESULTADOS OBTENIDOS/ESPERADOS

En la línea de investigación de recomendación de transporte multimodal, se espera que el sistema propuesto, mejore significativamente la precisión de las recomendaciones al considerar no solo factores como el tiempo de viaje y el costo, sino también las preferencias individuales y las condiciones del tráfico en tiempo real. Esto permitirá a los usuarios optimizar sus desplazamientos, reduciendo el tiempo de viaje y la congestión vehicular. Además, el sistema facilitará la integración de múltiples modos de transporte, lo que será especialmente útil en ciudades con infraestructuras complejas y alta densidad poblacional. La validación en diferentes entornos urbanos asegurará que el modelo sea adaptable a diversas condiciones, contribuyendo así a una movilidad más eficiente y sostenible.

En el ámbito de la identificación de residuos costeros, se espera que el sistema basado en técnicas de reconstrucción de imágenes con aprendizaje profundo logre una alta precisión en la detección y clasificación de residuos, superando el 90% de efectividad. Esto permitirá una gestión más eficiente de los desechos en áreas costeras, facilitando la implementación de políticas ambientales efectivas. El sistema también optimizará el uso del ancho de banda y el almacenamiento mediante técnicas de compresión de imágenes, lo que será crucial para su implementación en entornos remotos con limitaciones técnicas. La integración con tecnologías de edge computing y drones permitirá la captura y procesamiento de imágenes en tiempo real, mejorando la monitorización de áreas costeras de difícil acceso.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo se encuentra formado tres doctores en Ciencias de la Computación, un Magister en Ciencias de la Computación, tres maestrandos y un estudiante avanzado de grado de la UNPA.

Estas líneas de trabajo proporcionarán un marco propicio para la iniciación y/o finalización de estudios de posgrado de los integrantes docentes.

5. BIBLIOGRAFIA

Bochkovskiy, A.,Wang, C.-Y., & Liao, H.-Y. M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv preprint

Borrelle, S. B., Rochman, C. M., Liboiron, M., Bond, A. L., & Lavender Law, K. (2020). Predicted growth in plastic waste exceeds efforts to mitigate plastic pollution. Science, 369 (6510), 1515-1518.

Campigotto, P., Rudloff, C., Leodolter, M., & Bauer, D. (2016). Personalized and situation-aware multimodal route recommendations: the FAVOUR algorithm. *IEEE Transactions on Intelligent Transportation Systems*, *18*(1), 92-102.

Cheng, Z., Sun, H., Takeuchi, M., & Katto, J. (2020). Learned Image Compression with Discretized Gaussian Mixture Likelihoods and Attention Modules. https://arxiv.org/abs/2001.01568

Dai, J., Yang, B., Guo, C., & Ding, Z. (2015, April). Personalized route recommendation using big trajectory data. In *2015 IEEE 31st international conference on data engineering* (pp. 543-554). IEEE.

Fu, L., Sun, D., & Rilett, L. R. (2006). Heuristic shortest path algorithms for transportation applications: State of the art. *Computers & Operations Research*, *33*(11), 3324-3343.

Li, B., Liang, J., & Han, J. (2022). Variable-Rate Deep Image Compression With Vision Transformers. IEEE Access, 10, 50323-50334.

Liu, H., Han, J., Fu, Y., Li, Y., Chen, K., & Xiong, H. (2023). Unified route representation learning for multi-modal transportation recommendation with spatiotemporal pre-training. *The VLDB Journal*, *32*(2), 325-342.

Liu, H., Li, Y., Fu, Y., Mei, H., Zhou, J., Ma, X., & Xiong, H. (2020, August). Polestar: An intelligent, efficient and national-wide public transportation routing engine. In *Proc. of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 2321-2329).

Liu, H., Tong, Y., Zhang, P., Lu, X., Duan, J., & Xiong, H. (2019, July). Hydra: A personalized and context-aware multi-modal transportation recommendation system. In *Proc. of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 2314-2324).

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR),779-788.

Rogers, S., & Langley, P. (1998, January). Personalized driving route recommendations. In *Proc. of the American Association of Artificial Intelligence Workshop on Recommender Systems* (pp. 96-100).

Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. arXiv preprint arXiv:1804.02767

Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., & Monfardini, G. (2008). The graph neural network model. *IEEE transactions on neural networks*, *20*(1), 61-80.

Toli, A. M., & Murtagh, N. (2020). The concept of sustainability in smart city definitions. Frontiers in Built Environment, 6, 77.

United Nations Environment Programme - UNEP. (2023). Turning off the Tap: How the world can end plastic pollution and create a circular economy. UNEP. https://www.unep.org/resources/publication/turning-tap-how-world-can-end- plastic-pollution-and- createcircular-economy

Wang, J., Wu, N., Zhao, W. X., Peng, F., & Lin, X. (2019, July). Empowering A* search algorithms with neural networks for personalized route recommendation. In *Proc. of the 25th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 539-547).

Wang, C.-Y., Bochkovskiy, A., Liao, H.-Y. M., Yeh, I.-H.,Wu, S.-W., Chen, P.-Y., & Hsieh, H.-I. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. International Journal of Computer Vision, 131, 1273-1293.

Xu, A., Zhong, P., Kang, Y., Duan, J., Wang, A., Lu, M., & Shi, C. (2023). THAN: Multimodal Transportation Recommendation With Heterogeneous Graph Attention Networks. IEEE Transactions on Intelligent Transportation Systems.

Entrenamiento de GPT Personalizado para Telemetría en Tiempo Real

Daiana Jimena Ramírez [1], Matilde Inés Césari[1], Matilde Césari[1]; María Stefanoni[1]; Ana Muñoz[1]; Daniela Carbonari[1]; Bruno Roberti[1]; Santiago Pérez[1]

[1]GIDECo Grupo Regional UTN de Investigación y Desarrollo de Ecosistemas de Conocimiento - ReAVi Laboratorio de investigación Realidad Aumentada y Virtual) – Universidad Tecnológica Nacional FRM / matilde.cesari@frm.utn.edu.ar

RESUMEN

El proyecto consiste en el desarrollo de un asistente virtual para productores, diseñado para supervisar y optimizar el proceso de fermentación sin necesidad de intervención manual constante. Utilizará inteligencia artificial para analizar datos en tiempo real y sugerir acciones correctivas. Permitirá automatizar el control de equipos (ventiladores, estufas, refrigeración, humidificadores) según las condiciones del proceso. A través de una API, se conectará con ChatGPT y una base de datos, permitiendo consultas en lenguaje natural y generación de reportes personalizados. El objetivo es brindar a los pequeños productores una herramienta de monitoreo y control remoto, accesible desde cualquier lugar, sin la necesidad de contratar personal especializado

Palabras clave:. asistente virtual, monitoreo, ChatGPT

INTRODUCCIÓN

La ingeniería de prompts implica diseñar y ajustar prompts para optimizar las respuestas de modelos de lenguaje natural. Estudios como el de Si et al. (2022) han demostrado que prompts adecuados pueden mejorar significativamente la fiabilidad de GPT-3 en términos de generalización y reducción de sesgos. La integración de sistemas IoT con plataformas como ThingSpeak y Matlab ha sido explorada por Benyezza et al. (2018) y Verma et al. (2020), quienes demostraron la viabilidad de utilizar estos sistemas para la gestión de datos en tiempo real. Este proyecto proporcionará un caso de estudio sobre la ingeniería de prompts aplicada a un sistema de telemetría, ofreciendo una valiosa contribución a la personalización de modelos GPT en aplicaciones específicas. Además, explorará cómo la integración de GPT con APIs, como se describe en el trabajo de Amatriain (2024), puede mejorar el rendimiento del sistema

La ingeniería de prompts es esencial para el desarrollo de modelos GPT personalizados, permitiendo su aplicación en casos específicos. Este proyecto se centrará en el control de sistemas de telemetría de variables como la temperatura del alimento y las condiciones ambientales de la cámara (temperatura y humedad) durante procesos de fermentación sólida, que duran alrededor de 28 horas y se monitorearán en otoño/invierno. Utilizando datos recogidos por sensores conectados a microcontroladores Arduino y enviados a la plataforma ThingSpeak de Matlab, se evaluarán diversas técnicas de ingeniería de prompts aplicadas a GPT-4 para mejorar la precisión y eficiencia en el monitoreo y control del proceso.

Este proyecto se enmarca en el PID "Sistema remoto de supervisión y control con contenido de realidad aumentada, para pequeños emprendimientos en el área de alimentos funcionales" (Código del Proyecto: ASECME0008759TC), dirigido por la Dra. Matilde Inés Césari en el laboratorio ReAVi (Realidad Aumentada y Virtual) del Grupo GiDeCo (Grupo Regional UTN de Investigación y Desarrollo de Ecosistemas de Conocimiento) en la Universidad Tecnológica Nacional FRM

El proyecto busca desarrollar un **asistente virtual para productores**, que utilice inteligencia artificial y telemetría IoT para monitorear y controlar procesos de fermentación de manera remota.

Objetivo:

- Desarrollar un **asistente inteligente basado en GPT** que ayude a los productores a gestionar su proceso de fermentación sin supervisión constante.
- Integrar sensores y actuadores con la API de ThingSpeak para **automatizar acciones** como encender o apagar ventiladores, estufas, refrigeración o humidificadores.
- Conectar **ChatGPT con una base de datos** para permitir consultas en lenguaje natural y mejorar la toma de decisiones

El problema principal es la falta de personal en pequeños emprendimientos, lo que dificulta la supervisión continua de los procesos de fermentación y puede llevar a pérdidas o baja calidad en la producción.

Propuesta de solución:

- Asistente Virtual Inteligente: Un sistema basado en IA que actúa como un supervisor remoto, analizando datos y sugiriendo o ejecutando acciones automáticamente.
- Automatización de control: Se integrarán sensores y actuadores para que el sistema pueda encender o apagar dispositivos según las condiciones detectadas.
- Interacción en lenguaje natural: A través de ChatGPT y una base de datos conectada vía API, el usuario podrá consultar el estado del proceso y recibir recomendaciones.
- Monitoreo en tiempo real: Los datos serán procesados en ThingSpeak y Matlab, permitiendo su análisis y almacenamiento para futuras optimizaciones.

Oportunidades y contribución:

- Accesibilidad para pequeños productores: No requiere conocimientos avanzados en tecnología ni inversiones costosas.
- Automatización y eficiencia: Reduce errores humanos y optimiza el uso de recursos energéticos.
- Expansión a otras industrias: Puede adaptarse a otros procesos productivos que requieran monitoreo y control remoto

ESTADO DEL ARTE

El desarrollo de GPT personalizado mediante ingeniería de prompts ha sido objeto de diversas investigaciones. Si *et al.* (2022) detallan cómo los prompts adecuados pueden mejorar la fiabilidad de GPT-3 en términos de generalización y reducción de sesgos sociales. Kumar *et al.* (2022) investigan el diseño de prompts para optimizar chatbots basados en GPT-3 en aplicaciones específicas, como la gestión del bienestar mental, destacando la importancia de los prompts en la personalización y efectividad de los modelos GPT. Además, Nikitina (2023) proporciona técnicas detalladas sobre cómo diseñar y refinar prompts para obtener respuestas deseadas de modelos NLP. La integración de GPT con APIs es otro aspecto crucial, como se muestra en el trabajo de Amatriain (2024), quien introduce conceptos como "Chain-of-Thought" y "Reflection" en la construcción de agentes basados en LLM.

METODOLOGÍA

El proyecto utilizará una metodología mixta que combina el desarrollo de prompts con la integración de sistemas IoT. Inicialmente, se configurarán sensores conectados a microcontroladores Arduino para recoger datos de temperatura del alimento y condiciones ambientales de la cámara durante el otoño e invierno. Estos datos se enviarán a ThingSpeak de Matlab para su análisis. Posteriormente, se diseñarán y ajustarán prompts específicos para GPT-4, evaluando a través de métricas, su efectividad en el análisis y control de estos datos. La integración de datos se realizará mediante la API de ThingSpeak, permitiendo que la IA acceda y procese la información en tiempo real

CONCLUSIONES

Se espera que la investigación demuestre la efectividad de la ingeniería de prompts para personalizar modelos GPT en el control de sistemas de telemetría en tiempo real. Los resultados proporcionarán perspectivas valiosas sobre cómo optimizar las interacciones con modelos de lenguaje natural, mejorando la precisión y eficiencia en el monitoreo y control de estos procesos. Este proyecto no solo contribuirá al campo de la inteligencia artificial aplicada, sino que también ofrecerá una base sólida para futuras investigaciones en la personalización de modelos GPT en otras aplicaciones específicas.

REFERENCIAS

Benyezza, H., Bouhedda, M., Djellout, K., & Saidi, A. (2018). Smart Irrigation System Based on ThingSpeak and Arduino. 2018 International Conference on Applied Smart Systems (ICASS), 1-4.

Dang, C., Nguyen, K. D., & Dao, H. (2020). Apply Matlab in ThingSpeak Server to build the system measure and analyze data using IoT Gateway technology. Journal of Mechanical Engineering Sciences, 61, 88-95.

Kumar, H., Musabirov, I., Shi, J., Lauzon, A., Choy, K. K., Gross, O., Kulzhabayeva, D., & Williams, J. J. (2022). Exploring The Design of Prompts For Applying GPT-3 based Chatbots: A Mental Wellbeing Case Study on Mechanical Turk. ArXiv.

Si, C., Gan, Z., Yang, Z., Wang, S., Wang, J., Boyd-Graber, J. L., & Wang, L. (2022). Prompting GPT-3 To Be Reliable. ArXiv.

Verma, M., Kumar, V., & Sandeep. (2020). Implementation of an IoT based Real Time Signal Acquisition System over the Cloud using ThingSpeak Platform for Multi-Zonal Temperature & Humidity Measurement. International Journal of Research and Technology in Engineering and Research (IJRTER), 6, 1-7.

Nikitina, I. P. (2023). Prompt Engineering For ChatGPT: A Quick Guide To Techniques, Tips, And Best Practices.

Amatriain, X. (2024). Prompt Design and Engineering: Introduction and Advanced Methods. ArXiv.

ESTRATEGIAS Y ALGORITMOS DE OPTIMIZACIÓN MEDIANTE APRENDIZAJE AUTOMÁTICO APLICADAS A SISTEMAS BASADOS EN ENERGÍAS RENOVABLES

Marcelo Cappelletti[1,2], Lucas Olivera[1,3], Harry Saltos[2,3], Carlos Pinzón Rueda[4], Jésica Guzmán[1], Carlos Schenone[1], Jorge Osio[1,2], Fernando Alvira[4], Arles Gil Rebaza[5,6], Nicole Denon[1], Christian Botta[1], Patricio Gross[1], Leonardo Amet[1], Martín Morales[1,7]

[1] Programa TICAPPS, Univ. Nac. Arturo Jauretche, Florencio Varela (1888), Argentina.
[2] Grupo de Control Aplicado, Instituto LEICI (UNLP-CONICET), La Plata (1900), Argentina.
[3] Becario de la Comisión de Invest. Cient., Pcia. de Bs. As. (CICPBA), La Plata, (1900), Argentina.
[4] Laboratorio SiCoBioNa, Univ. Nac. Quilmes (UNQ), Bernal, Argentina.
[5] Dpto. de Física, Fac. de Ciencias Exactas, Univ. Nac. La Plata, La Plata (1900), Argentina.
[6] Instituto de Física La Plata (IFLP, UNLP-CONICET), La Plata (1900), Argentina.
[7] Centro UTN CODAPLI-FRLP, La Plata (1900), Argentina.
e-mail: mcappelletti@unaj.edu.ar

RESUMEN

El aumento del consumo energético promueve el uso de nuevas tecnologías basadas en recursos renovables para generar electricidad en detrimento de los combustibles fósiles (carbón, petróleo, gas natural), lo cual a su vez contribuye a reducir las emisiones de dióxido de carbono a la atmósfera y proteger nuestro medio ambiente.

En particular, la energía solar fotovoltaica constituye una de las opciones más prometedoras, dado que se presenta como una fuente limpia y renovable de electricidad con bajo impacto ambiental, disponibilidad en el sitio de consumo y sostenibilidad. Sin embargo, existen ciertas dificultades para integrar esta energía a la red eléctrica. Las razones son principalmente la variabilidad del recurso solar, incluyendo la falta de predicción del mismo, las dificultades para reemplazar a los generadores tradicionales en algunos servicios de red, y la importante inversión inicial requerida para la instalación de plantas de generación basadas en conversión fotovoltaica.

El objetivo general del presente proyecto es el de contribuir al desarrollo sostenible realizando aportes científico-tecnológicos que redunden en métodos y procesos más eficientes de generación de energía eléctrica a partir de recursos energéticos renovables como lo es la energía solar. Estas contribuciones se darán tanto en el diseño de celdas solares de nuevos materiales, como así también en el análisis de sistemas fotovoltaicos (SFV), con el fin de reducir los costos asociados a la producción energética para que estos emprendimientos resulten competitivos y puedan ser aplicables a nivel local y regional, tanto para actividades domésticas como industriales.

En este proyecto, la investigación se enfocará sobre estrategias y algoritmos de optimización mediante el uso de aprendizaje automático (machine learning), aplicadas a dos ejes temáticos diferentes, por un lado, se apuntará al estudio de dispositivos de conversión fotovoltaica basados en nuevos materiales, y, por otro lado, se estudiarán SFV bajo diversas condiciones de cielo.

Asimismo, el proyecto propicia la formación de recursos humanos, tanto de docentes investigadores como de estudiantes de grado y posgrado.

Finalmente, la investigación propuesta brinda la posibilidad de continuar y fortalecer el trabajo en conjunto con otros centros académicos del área de influencia de la UNAJ abocados al tema en cuestión.

Palabras clave: *Aprendizaje automático, optimización, energía solar.*

CONTEXTO

Las líneas de I/D presentadas en este trabajo se encuentran en el marco del Programa de Investigación TICAPPS (Tecnologías de la Información y la Comunicación en Aplicaciones de Interés Social), UNAJ (Res. N° 002/24), bajo la Dirección del Dr. Ing. Morales, y actualmente es financiada por los proyectos de investigación: UNAJ INVESTIGA 2023 (80020230100011UJ), PIP CONICET 2021–2023 (112-202001-01460CO) y PICTO-UNAJ (2022-02-00008), todos bajo la Dirección del Dr. Ing. Cappelletti.

1. INTRODUCCIÓN

En la actualidad, se entiende que un sistema es inteligente cuando tiene la capacidad de adaptarse y generar respuestas adecuadas ante su entorno. El término Inteligencia Artificial (IA) hace referencia a cualquier técnica que permita a las computadoras replicar aspectos de la inteligencia humana mediante el uso de expresiones lógicas y representaciones abstractas [1]. Estas técnicas pueden ser empleadas para modelar, identificar, optimizar, predecir y controlar el comportamiento dinámico de diversos sistemas en el ámbito real.

El Aprendizaje Automático (Machine Learning) constituye una rama de la IA que emplea diversos algoritmos capaces de aprender de su entorno, con el objetivo de identificar patrones relevantes dentro de un conjunto de datos. Este aprendizaje se basa en la información proporcionada durante la etapa de entrenamiento del algoritmo [2]. El término 'aprendizaje' en el contexto de las computadoras hace referencia a la capacidad para reconocer patrones en grandes volúmenes de datos, lo que les permite tomar decisiones o realizar predicciones sobre comportamientos futuros de un sistema determinado.

En los últimos años, el aprendizaje automático ha adquirido gran relevancia como herramienta en una amplia variedad de tareas que implican la extracción de información a partir de grandes cantidades de datos. Actualmente, la tecnología basada en aprendizaje automático está presente en numerosos ámbitos, como en medicina, seguridad, turismo, finanzas, producción energética, robótica, entre otras. Algunos ejemplos de dichos avances son: diagnóstico de enfermedades, conducción autónoma, recomendaciones personalizadas de videos y búsquedas en Internet, reconocimiento de voz y facial, entre otras.

Este trabajo se centra en algoritmos de aprendizaje automático supervisado, en los cuales los modelos son entrenados utilizando un conjunto de datos en el que la respuesta correcta es conocida. En este enfoque, se presentan a los modelos patrones de entrada junto con sus respectivas salidas esperadas, y los hiperparámetros del modelo se ajustan para minimizar el error entre la salida real y la salida esperada.

Estos algoritmos pueden aplicarse tanto en modelos de regresión como de clasificación, dependiendo del tipo de resultado deseado. Los modelos de regresión son modelos matemáticos cuyo objetivo es identificar la relación entre una variable dependiente y una o más variables explicativas o independientes.

En estos modelos, el resultado es un valor numérico, perteneciente a un conjunto infinito de posibles valores. En contraste, los modelos de clasificación asignan un resultado a una clase dentro de un número limitado de clases, según el problema en cuestión, por ejemplo: sí-no, aceptable-no aceptable, entre otros.

A continuación, se presenta un breve marco teórico de cada una de las dos áreas de estudio investigadas.

Dispositivos de conversión fotovoltaica basados en nuevos materiales:

Las celdas solares son dispositivos fotovoltaicos que convierten la radiación solar en energía eléctrica, con un reducido costo operativo y bajos niveles de contaminación [3]. Estos dispositivos constituyen los componentes principales de los SFV. El

diseño y optimización de celdas solares, es extremadamente importante hoy en día en el campo de las energías alternativas, dado que uno de los problemas que se presenta es que, de toda la energía recibida por la celda solar, solamente se aprovecha una pequeña fracción, es decir que la relación de transferencia es todavía baja, por lo que pensar en cómo mejorar sus rendimientos es una tarea que despierta gran interés. Actualmente, cerca del 90% de las celdas solares disponibles en el mercado mundial están basadas en silicio mono- y poli-cristalino, la cual si bien es una tecnología probada, confiable y de larga vida útil, su eficiencia ha alcanzado valores máximos cercanos a su límite teórico.

El estudio de las propiedades microscópicas de los materiales, junto con las técnicas de modelado y simulación, son herramientas fundamentales para predecir y analizar el comportamiento de los dispositivos. Los avances en los procesos de fabricación de los dispositivos fotovoltaicos, ha proporcionado un camino para desarrollar nuevas tecnologías en celdas solares con aceptables eficiencias, menor tiempo de procesamiento y bajo costo.

Uno de los desarrollos que ha ganado especial relevancia en los últimos años son las celdas solares de perovskita (CSP), que han surgido como una tecnología con potencial para revolucionar la industria fotovoltaica. Esto se debe principalmente a sus excelentes propiedades ópticas y electrónicas [4, 5], como banda prohibida sintonizable, gran coeficiente de absorción, altas movilidades de los portadores de carga y largas longitudes de difusión, y su proceso de fabricación más simple y menor costo en comparación con las celdas solares de silicio cristalino convencional.

Estos dispositivos han incrementado de manera significativa su eficiencia desde su aparición: desde 3.8% en el año 2009 [6] hasta 26% en el año 2023 [7]. Sin embargo, a pesar del alto rendimiento proporcionado por las CSP, desafortunadamente estos compuestos son sensibles a cambios de humedad. Además, la presencia de plomo (Pb) lo vuelve tóxico y poco amigable al medioambiente [8]. Estos factores podrían obstaculizar la futura comercialización de estos dispositivos, por lo que mayores estudios son necesarios para remediar estos inconvenientes.

Sistemas fotovoltaicos bajo diversas condiciones de cielo:

El conocimiento preciso de la radiación solar en un determinado instante y lugar es un factor clave no solamente para el dimensionamiento óptimo de los sistemas fotovoltaicos (SFV) para la generación de electricidad, sino también en los colectores solares para el calentamiento de fluidos, en el diseño arquitectónico y en la agricultura, entre otras áreas.

En particular, uno de los inconvenientes que presentan los SFV para maximizar la conversión de energía es la variabilidad del recurso solar, incluyendo la falta de predicción del mismo.

El problema que se presenta es que no siempre es posible contar con datos experimentales de la radiación solar en los lugares de interés [9]. Además, la radiación solar disponible en un punto determinado de la superficie terrestre, es fuertemente dependiente de la localización geográfica (latitud, longitud, altura sobre el nivel del mar) y de los microclimas locales del sitio (temperatura, humedad, entre otras variables meteorológicas), con lo cual se trata de una magnitud variable con el tiempo de manera instantánea, horaria, diaria y estacional.

Adicionalmente, la cobertura de nubes es un factor que atenúa y ocasiona intermitencia en los valores de la energía proveniente del sol. Esta variabilidad de la radiación solar provoca que la producción de energía eléctrica que pueden generar los SFV no sea constante ni totalmente predecible en el tiempo, lo cual es una desventaja con respecto a los combustibles fósiles. El conocimiento preciso de la radiación solar que alcanza un SFV en un determinado instante y lugar tiene un rol determinante para: (i) la evaluación de la producción potencial de energía eléctrica que podría obtenerse en ese lugar, ofreciendo

una mayor previsibilidad a este tipo de tecnología, para que pueda satisfacer la demanda en los momentos de intermitencia; (ii) el desarrollo de estrategias que permitan operar al SFV en el punto de máxima potencia la mayor parte del tiempo.

Si bien la radiación solar puede ser medida por piranómetros y pirheliómetros, se trata de dispositivos costosos que requieren ser calibrados a intervalos regulares y con un apropiado mantenimiento para asegurar la calidad de los datos [10]. Por este motivo, se han utilizado métodos teóricos para su determinación, entre ellos, modelos lineales, polinómicos, exponenciales y logarítmicos, técnicas geoestadísticas, técnicas estocásticas y técnicas basadas en imágenes satelitales [11, 12]. Sin embargo, los modelos mencionados pueden sufrir problemas de precisión debido a la incertidumbre de las condiciones iniciales, como consecuencia de la cobertura de nubes.

Gracias al crecimiento de las capacidades de cálculo y al mejoramiento de los algoritmos implementados, en los últimos años se han comenzado a utilizar herramientas basadas en aprendizaje automático, como una buena alternativa para la estimación de la radiación solar con el mínimo error [13-15].

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Las líneas de I/D que se presentan en este trabajo están basadas en el estudio y desarrollo de herramientas de aprendizaje automático para el diseño, producción y aprovechamiento de energía eléctrica a partir de recursos energéticos renovables, específicamente de la energía solar fotovoltaica, con el propósito de reducir los costos asociados, para que estos emprendimientos resulten competitivos.

Se espera que las herramientas utilizadas presenten alto rendimiento en la precisión de los resultados, bajo costo de cómputo, facilidad de implementación y buena velocidad de convergencia.

Dichas herramientas serán aplicadas para:

(a) El diseño y optimización de materiales y dispositivos fotovoltaicos de bajo costo.

(b) La optimización de SFV para diferentes condiciones de cielo.

Ambas líneas de I/D se abordarán de manera teórica/experimental. El estudio conjunto de materiales, dispositivos y sistemas basados en energías renovables mediante técnicas de aprendizaje automático es innovador y atractivo, lo cual favorece el hecho de que estudiantes, graduados o investigadores, con interés en estos temas, puedan participar.

3. RESULTADOS ESPERADOS

Con respecto a la línea de I/D relacionada con el diseño y optimización de materiales y dispositivos fotovoltaicos de bajo costo, se espera contribuir al desarrollo de una tecnología prometedora en el ámbito de los dispositivos fotovoltaicos, como son las celdas solares de perovskita, con el fin de mejorar su eficiencia y abordar los problemas de estabilidad térmica y toxicidad, con miras a su futura comercialización.

Mientras que con respecto a la línea de I/D relacionada con la optimización de SFV para diferentes condiciones de cielo, se espera poder brindar herramientas alternativas basadas en aprendizaje automático, para determinar la atenuación que sufre la radiación solar bajo diversas condiciones de cielo, con el fin de adaptar la producción de electricidad a la presencia de nubes, logrando así una optimización de la respuesta de los SFV y una reducción de costos de esta tecnología. Asimismo, esto será beneficioso para diversas aplicaciones, entre ellas para la actividad agroindustrial, en pos de promover el uso de energías alternativas en el territorio de la UNAJ.

4. FORMACIÓN DE RECURSOS HUMANOS

Uno de los principales objetivos del Programa TICAPPS, dentro de las líneas de I/D presentadas en este trabajo, es la formación de recursos humanos, tanto docentes investigadores como estudiantes, especializados en el campo de la Inteligencia Artificial, particularmente en herramientas basadas en aprendizaje automático. El programa busca fomentar el desarrollo de soluciones innovadoras y de alta calidad, lo cual resulta fundamental para el avance tecnológico del territorio de la UNAJ.

El equipo de trabajo de las líneas de I/D está conformado por siete docentes investigadores formados (seis Doctores y un Magister), cinco docentes investigadores en formación, y dos estudiantes avanzados, todos autores de este artículo. Actualmente, se encuentran en curso cuatro Tesis de Doctorado, todas ellas relacionadas directamente con las líneas de I/D presentadas.

Finalmente, la investigación propuesta ofrece la oportunidad de continuar y fortalecer la colaboración con otros centros académicos de la región de influencia de la UNAJ, dedicados al tema en cuestión. Entre estos se incluyen el Laboratorio de Simulación Control Biofotónica y Nanotecnología (SiCoBioNa) del Departamento de Ciencia y Tecnología de la Universidad Nacional de Quilmes (UNQ), el Grupo de Control Aplicado (GCA) del Instituto LEICI (UNLP-CONICET), y el Instituto de Física La Plata (IFLP-UNLP-CONICET).

5. BIBLIOGRAFÍA

[1] N. Nilsson. "Principles of Artificial Intelligence", Springer-Verlag, New York (1980).

[2] E. Alpaydin. "Machine Learning: The New AI". Editorial MIT Press. (2016).

[3] K. Taretto. "Celdas Solares, Teoría, Ensayo y Diseño". Ed. Nueva Librería. Ed. 1º. 2015.

[4] J. Burshka, et al. "Sequential deposition as a route to high-performance perovskite-sensitized solar cells". Nature. Vol. 499, pp.316-319 (2013). DOI: https://doi.org/10.1038/nature12340

[5] V. Gonzalez-Pedro et al.. "General working principles of $CH_3NH_3PbX_3$ perovskite solar cells", Nano Lett. Vol. 14, pp 888-893 (2014). DOI: https://doi.org/10.1021/nl404252e

[6] A. Kojima, K. Teshima, Y. Shirai y T. Miyasaka. "Organometal halide perovskites as visible-light sensitizers for photovoltaic cells". J. Am. Chem. Soc. Vol. 131 (17), pp.6050-6051, (2009). DOI: 10.1021/ja809598r

[7] NREL, 2023. Best Research-Cell Efficiency Chart (https://www.nrel.gov/pv/cell-efficiency.html)

[8] W. Ke y M. Kanatzidis. "Prospects for low-toxicity lead-free perovskite solar cells". Nature Commun., Vol. 10, pp.965-968, (2019). DOI: doi.org/10.1038/s41467-019-08918-3

[9] C. Raichijk, H. Grossi Gallegos, R. Righini. "Evaluación de un método alternativo para la estimación de valores medios mensuales de irradiación global en Argentina". Avances en Energías Renovables y Medio Ambiente, Vol. 9. pp. 11.05–11.08, (2005).

[10] C. Pandey, K. Katiyar. "Solar Radiation: Models and Measurement Techniques". Journal of Energy. Vol. 2013, Article ID 305207, 8 pages, (2013). DOI: doi.org/10.1155/2013/305207

[11] A. Assi, M. Jama, M. Al-Shamisi. "Prediction of Global Solar Radiation in Abu Dhabi". Int. Scholarly Research Network ISRN Renewable Energy. Vol. 2012, Article ID 328237, 10 pages, (2012). DOI: doi.org/10.5402/2012/328237

[12] J. Polo Martínez. "Optimización de modelos de estimación de la radiación solar a partir de imágenes de satélite". Tesis Doctoral. Universidad Complutense de Madrid, (2010).

[13] Z. Pang, F. Niu, and Z. O'Neill. "Solar radiation prediction using recurrent neural network and artificial neural network: A case study with comparisons". Renewable Energy, 156:279–289, 2020.

[14] P. Kumari and D. Toshniwal. "Extreme gradient boosting and deep neural network based ensemble learning approach to forecast hourly solar irradiance". Journal of Cleaner Production, 279:123285, 2021.

[15] M. Mukhtar, et al. "Development and comparison of two novel hybrid neural network models for hourly solar radiation prediction". Applied Sciences, 12(3), 2022.

Exploración de ROS (Robot Operating System) para el uso en vehículos no tripulados de superficie

Zaradnik, Ignacio; Turconi, Diego; Lupi, O.Daniel. Caccaviello, Diego.
Laboratorio de Inteligencia Ambiental Departamento de Ingeniería e Investigación Tecnológica, Universidad Nacional de La Matanza. Buenos Aires, Argentina
izaradnik@unlam.edu.ar;dturconi@unlam.edu.ar;olupi@unlam.edu.ar;dcaccaviello@unlam.edu.ar

RESUMEN

Distintos avances tecnológicos van convergiendo hacia una mayor integración de la robótica a nuestras vidas, siendo un ejemplo el desarrollo de vehículos no tripulados de superficie para el análisis de ecosistemas acuáticos. La nanotecnología, la inteligencia artificial, sensores y actuadores avanzados y los nuevos materiales son elementos de este nuevo auge. Pero para amalgamar todos ellos en una aplicación específica, como la navegación sin colisiones y en un entorno no estructurado, es necesario el desarrollo de un software flexible y confiable. En base a esto se estudiaron las ventajas de la implementación de ROS (Robot Operating System), el cual es un software libre bajo licencia BSD, versus un desarrollo ad hoc. Identificadas las ventajas se decidió comenzar a trabajar con la versión 1 de ROS(ROS1), planteándose una arquitectura formada por un nodo central y uno remoto, e implementándose una comunicación entre ellos. El existo logró permitió familiarizarse con los distintos elementos del sistema ROS, comprender su arquitectura, adquirir experiencia en su uso y proyectar su uso en etapas futuras del proyecto.

Palabras Clave: Vehículo no tripulado de superficie (USV), Sistema operativo robótico (ROS), Nodos, Tópicos.

CONTEXTO

En el Laboratorio de Inteligencia Ambiental del Departamento de Ingeniería e Investigación Tecnológica de la Universidad Nacional de La Matanza, se está trabajando desde hace algunos años en aplicaciones para el monitoreo de ecosistemas acuáticos [1][2]. El presente trabajo es parte de lo realizado en el marco del proyecto "Utilización de inteligencia artificial en vehículos no tripulados de superficie para aplicaciones medioambientales", el cual se está desarrollado desde comienzos del 2024 y finalizara a fines del 2025. Este trabajo está financiado con fondos provenientes del Programa de Incentivos para Docentes Investigadores de la Secretaría de Políticas Universitarias (PROINCE).

1. INTRODUCCIÓN

Día a día la robótica se va empleando en un mayor número de campos, desde vehículos autónomos para hacer entregas, pasando por robots bípedos para asistencia en fábricas y depósitos, hasta robots con forma de animal para apoyo emocional, estos y muchos otros ejemplos se pueden encontrar en el portal de noticias sobre robots de "La Nación" [3]. El cuidado del medio ambiente, y en particular los ecosistemas acuáticos, no es una escisión, en [4] se presenta el control coordinado de vehículos aéreos y acuáticos para la navegación libre de colisiones en este tipo de aplicaciones. Esta mayor adopción se debe a los distintos avances tecnológicos que se han sucedido sucediendo: la mayor escala de integración de los procesadores que permite la implementación de arquitecturas paralelas de procesamiento [5]; la inteligencia artificial (IA) que permite un mejor desempeño por parte de los robots en áreas como la navegación y mapeo, representación de conocimiento y razonamiento, planificación, interacción y aprendizaje [6]; nuevos sensores y actuadores que mejoran la percepción del entorno que los rodea [7] y nuevos materiales que amplían sus habilidades y los medioambientes en los cuales se pueden emplear [8].

Como consecuencia de que la investigación y los desarrollos enfocados en la robótica se ha acelerado, en particular la necesidad de moverse

y actuar en entornos que nunca habían visto antes o descriptos semánticamente, es necesario contar una plataforma de un software flexible, escalable, confiable y reutilizable. Características las cuales se pueden encontrar en los softwares intermedios o middlewares. En robótica han existen varios: YARP (Yet Another Robot Platform), OROCOS (Open Robot Control Software), MOOS (Mission Oriented Operating Suite), ROS (Robot Operating System), entre otros. Cada uno de estos presentan características relevantes para determinado tipo de aplicación, tal como se detalla en [9], siendo ROS uno de los más difundidos actualmente.

ROS presenta una arquitectura distribuida constituida por nodos, uno central que coordina la comunicación entre los nodos distribuidos y estos últimos que envían y reciben información. Los nodos pueden ser desde tareas ejecutándose en la misma plataforma en que se encuentra el nodo central, hasta hardware tan disimile como un Arduino UNO ejecutando un sketch para el manejo motores o una placa NVIDIA corriendo algoritmos de SLAM (localización y mapeo simultáneos). La comunicación entre estos es bidireccional y se puede realizar por varios elementos (Tópicos, Servicios y Mensajes) cada uno con características especiales adecuados para la optimización del programa. Esta arquitectura permite a los programadores desarrollar un sistema complejo conectando soluciones existentes para problemas de menor escala.

2. LINEAS DE INVESTIGACIÓN y DESARROLLO

Los objetivos de este trabajo fueron familiarizarse con los distintos elementos del sistema ROS, comprender su arquitectura, adquirir experiencia en la instalación de las herramientas asociadas y en su uso, identificar las limitaciones y visualizar el potencial futuro. Para alcanzar estos objetivos se planteó: la instalación de ROS, tanto en PC como en sistemas embebidos; la implementación de una arquitectura distribuida de nodos y la comunicación bidireccional entre ellos por medio de un puerto serial.

2.1. Selección de la versión ROS

Al momento de realizar el presente trabajo, febrero del 2025, en la página de ROS (https://www.ros.org/) se recomiendan dos distribuciones de la versión 2 de ROS (ROS2), Jazzy Jalisco y Humble Hawksbill. Dichas distribuciones están asociadas a versiones de Ubuntu, versión 24.04 y 22.04 respectivamente. Independiente de estas recomendaciones se puede acceder a las distribuciones previas, tanto para la versión 1 como 2 de ROS. Para el presente proyecto se decidió trabajar con la distribución Kinetic Kame correspondiente a la versión 1 de ROS (ROS1), ya que la misma es la dirigida a la versión 16.04 de Ubuntu, embebida en la plataforma de hardware contemplada. La cual se basa en un procesador AllWinner R40 [10] de 4 núcleos ARM A7 con una frecuencia máxima de operación de 576Mhz.

2.2. Rosserial

El objetivo de la exploración de ROS es su futura utilización en un USV, cuyo sistema electrónico implementa una arquitectura distribuida sobre enlaces RS232 [2]. Como la comunicación en ROS esta inicialmente pensada para trabajar sobre una red Ethernet fue necesario hallar un elemento que permitiera que los nodos distribuidos del USV se comunicaran con el nodo central por medio de un enlace previsto. El resultado de esta búsqueda fue rosserial (https://wiki.ros.org/rosserial), el cual es un protocolo para envolver mensajes serializados ROS estándar y multiplexar múltiples temas y servicios a través de un dispositivo de caracteres como un puerto serie.

2.3. Instalaciones y pruebas iniciales

A continuación, se detalla la instalación de ROS y Rosserial.

1) Se configuró de la lista de fuentes para descargas. Se incluyó la URL de los paquetes de ROS para la versión 16-04 LTS de Ubuntu.

sudo sh -c 'echo "deb http://packages.ros.org/ros/ubuntu xenial main" > /etc/apt/sources.list.d/ros-latest.list'

2) Se descargó y generó la clave de autentificación del paquete ROS.

wget http://packages.ros.org/ros.key -O - | sudo apt-key add -

3) Se actualizó la lista de servidores de descarga para que se agregue la URL de ROS previamente configurada.

sudo apt-get update
4) Se descargo el paquete de ROS kinetic.
sudo apt-get install ros-kinetic-desktop
En caso de trabajar con una PC se puede instalar el paquete full.
sudo apt-get install ros-kinetic-desktop-full
5) Se cambió el modo ejecución del archivo de la instalación encargado de configurar las variables de entorno.
chmod +x /opt/ros/kinetic/setup.bash
6) Se inicializaron las dependencias para compilar componentes.
sudo rosdep init
rosdep update
7) Se configuraron las fuentes para ejecución de ROS y su ambiente.
echo "source /opt/ros/kinetic/setup.bash" >> ~/.bashrc
source ~/.bashrc
8) Se instaló el IDE de Arduino
sudo apt-get install arduino
9) Se instaló el rosserial
sudo apt update
sudo apt install ros-kinetic-rosserial-arduino ros-kinetic-rosserial
10) Se instaló la librería rosserial en Arduino
rosrun rosserial_arduino make_libraries.py ~/sketchbook/libraries
11) Se abrió el IDE de Arduino, se escribió el código que se encuentra a continuación, se compilo y descargo.

```
#include <ros.h>
#include <std_msgs/String.h>
ros::NodeHandle nh;
std_msgs::String str_msg;
ros::Publisher chatter("chatter", &str_msg);
char hello[] = "Hello from Arduino!";
void setup() {
  nh.initNode();
  nh.advertise(chatter);
}
void loop() {
  str_msg.data = hello;
  chatter.publish(&str_msg);
  nh.spinOnce();
  delay(1000);  // Send message every second
}
```

En dicho código se comienza incluyendo las librerías que proporcionan las funcionalidades necesarias para que el Arduino se comunique con ROS y maneje cadena de caracteres. A continuación, se generan tres objetos: un *NodeHandler*, para manejar la comunicación con ROS; un *String,* que contendrá el mensaje a enviar y un *Publisher,* que se encarga de publicar los mensajes asociándolos con el tópico *chatter*. Luego se define la cadena de caracteres a transmitir, se inicializa el nodo (*nh.initNode()*) y se registra el tópico en el cual se publicara (*nh.advertise(chatter)*). En el bucle principal se publica el mensaje periódicamente. En este, la función *nh.spinOnce()* gestiona la comunicación con ROS.
12) Se abrió una consola y disparo el ROS
roscore
13) Se abrió otra consola y conectó la placa Arduino. *ttyUSB* es el puerto donde se encuentra el Arduino Uno montado.
rosrun rosserial_python serial_node.py /dev/ttyUSB0
14) Se abrió una tercera consola para ver los mensajes publicados en el tópico *chatter*.
rostopic echo /chatter
A continuación, se puede observar los mensajes recibidos.
data: "Hello from Arduino!"

2.4. Comunicación bidireccional

Para esta prueba se escribió un nuevo programa para el Arduino y un script de Pyton para el nodo central. En el código de Arduino, además de los tres objetos previamente incluidos, se incluye un objeto *Subscriber* para escuchar el tópico */toggle.* A continuación, se encuentran definiciones, variables y la función que se ejecuta si el tópico */toggle* recibe un mensaje (*toggleCallback)*. Si el mensaje es *toggle*, se pone en "1" un flag. Luego se inicializa el nodo, se registra el tópico *nodo_pub* y se suscribe al tópico *toggle_sub (nh.subscribe(toggle_sub)).* En el bucle principal se llama a la función de gestión de ROS, se publica en nombre del nodo y se chequea si el flag de recepción de *toggle* esta activado. En caso afirmativo se hace parpadear el led durante 5 segundos y luego se apaga hasta que se reciba otro mensaje.

```
#include <ros.h>
#include <std_msgs/String.h>
ros::NodeHandle nh;
std_msgs::String nodo_msg;
ros::Publisher nodo_pub("/nodo_arduino", &nodo_msg);
ros::Subscriber<std_msgs::String> toggle_sub("/toggle", &toggleCallback);

#define LED_PIN 13
bool toggleActive = false;
unsigned long toggleStartTime = 0;
void toggleCallback(const std_msgs::String &msg) {
  if (strcmp(msg.data, "toggle") == 0) {
    toggleActive = true;
    toggleStartTime = millis();
    Serial.println("Recibido: toggle → LED parpadeará");}}

void setup() {
  pinMode(LED_PIN, OUTPUT);
  Serial.begin(115200);
  nh.initNode();
  nh.advertise(nodo_pub);
  nh.subscribe(toggle_sub);
  nodo_msg.data = "Arduino_LED_Node";
}
void loop() {
  nh.spinOnce();
  nodo_pub.publish(&nodo_msg);

  if (toggleActive) {
    unsigned long elapsed = millis() - toggleStartTime;
    if (elapsed < 5000) { // Parpadeo por 5 segundos
      digitalWrite(LED_PIN, (millis() / 500) % 2); // LED parpadea a 1 Hz
    } else {
      toggleActive = false;
      digitalWrite(LED_PIN, LOW); // Apagar LED después de 5s
    }
  } else {
    digitalWrite(LED_PIN, LOW); // LED OFF
  }
  delay(100);
}
```

En el script de Pyton, se comienza importando la librería de ROS y de manejo de cadenas. A continuación, se crea la clase *LedController.* El constructor de la misma inicializa el nodo (*rospy.init_node("led_publisher")*), genera un objeto de ROS para publicar en el tópico */toggle* y otro para subscribirse al tópico */nodo_arduino* y define variables a usar. A continuación, se definen los métodos asociados a la clase: *callback_arduino(self, msg)* y *run(self).* El primer método se ejecuta si se recibe un mensaje en el tópico subscripto, y pondrá un "1" en un flag si en el mensaje es *Arduino_LED_Node.* El segundo método se ejecuta siempre que ROS este activado (*while not rospy.is_shutdown())* y cheque el flag que maneja el método *callback_arduino*, si tiene un "1" publica en el tópico */toggle* el mensaje *toggle*. En el bloque principal se crea una instancia de *LedController* y se ejecuta el método *run.*

```
import rospy
from std_msgs.msg import String

class LedController:
    def __init__(self):
        rospy.init_node("led_publisher")
        self.pub = rospy.Publisher("/toggle", String, queue_size=10)
        rospy.Subscriber("/nodo_arduino", String, self.callback_arduino)
        self.rate = rospy.Rate(0.5)
        self.arduino_detectado = False

    def callback_arduino(self, msg):
        if msg.data == "Arduino_LED_Node":
            rospy.loginfo("Arduino detectado correctamente. Enviando 'toggle'.")
            self.arduino_detectado = True
        else:
            rospy.logwarn("Nombre de nodo desconocido: %s", msg.data)
            self.arduino_detectado = False

    def run(self):
        while not rospy.is_shutdown():
            if self.arduino_detectado:
                self.pub.publish("toggle")
                rospy.loginfo("Enviado: toggle")
            self.rate.sleep()
```

```
if __name__ == "__main__":
    controller = LedController()
    controller.run()
```

La ejecución de este programa se realiza a través del siguiente comando en el paso 14.

rosrun tu_paquete led_publisher.py

3.RESULTADOS OBTENIDOS

3.1. Resultados

A través del primer ejemplo se presenta el proceso de instalación de ROS en PC y en sistemas embebidos, así como se ejemplifica el uso del elemento Tópico. Con el segundo ejemplo se presenta una comunicación bidireccional entre dos elementos de un sistema con arquitectura distribuida.

3.2. Conclusiones

Se logra comprender la arquitectura de funcionamiento de ROS y conocer sus distintos elementos. Se adquirió experiencia en su instalación y en el uso de los tópicos como elemento de comunicación bidireccional. En lo que respecta a los alcances de este middleware se aprecia un potencial muy grande ya que se pueden incorporar paquetes de SLAM Google o ROS2 Cartographer, los cuales facilitarían la navegación del USV. Como siguiente paso se implementará un sistema de evasión de obstáculos, el cual estará constituido por tres nodos: el central, el encargado de detectar los obstáculos (por medio de ultrasonido y Lidar) y el encargado del control de los motores.

4. FORMACIÓN DE RECURSOS HUMANOS

El ámbito de este proyecto permitió tanto la formación grupal del equipo de trabajo, así como la individual de cada uno de sus miembros. La formación grupal buscó generar conocimiento en la temática de midlewares para robótica con foco en ROS. En relación las formaciones individuales, los autores lograron los objetivos planteados en la línea de investigación y desarrollo.

5. REFERENCIAS

[1] Lupi; Zaradnik; Turconi; Slawiski; Caccaviello, “Miniboyas ambientales para monitoreo de ecosistemas acuáticos” en el CASE 2023. ISBN: 978-631-90145-0-1.

[2] Zaradnik; Lupi; Salvatore; Caccaviello, “Prototipo de vehículo no tripulado de superficie para aplicaciones ambientales”. Cuarto congreso virtual de microcontroladores y sus aplicaciones. ISBN:978-950-42-0246-2.

[3] LA NACION. “Robots”. Accedido el 16/02/2025.[Online]. Disponible: https://www.lanacion.com.ar/tema/robots-tid48565/

[4] L. Garberoglio, P. Moreno, I. Mas and J. I. Giribet, "Coordinated asv-uav control for marine collision-free navigation", *2019 XVII Workshop on Information Processing and Control (RPIC)*, pp. 1-6, 2019.

[5] M. R. Islam, "Parallel Computing Architectures for Robotic Applications: A Comprehensive Review," *arXiv preprint*, arXiv:2407.01011, Jul. 2024. [Online]. Disponible: https://arxiv.org/abs/2407.01011.

[6] Kunze, N. et al., “Artificial intelligence for long-term robot autonomy: A survey,” IEEE Robot. Autom. Lett., vol. 3, no. 4, pp. 4023–4030, Oct. 2018, doi: 10.1109/LRA.2018.2860628.

[7] E. Ackerman. “Robots with warm skin know what they're touching”. IEEE Spectrum. Accedido el 15/02/2025. [Online].Disponible: https://spectrum.ieee.org/robots-with-warm-skin-know-what-theyre-touching

[8] A. Cuthbertson. “Ciencia: un nuevo material que cambia de forma podría revolucionar la robótica”. Independent español. Accedido el 16/02/2025. [Online]. Disponible: https://www.independentespanol.com/noticias/ciencia/ciencia-material-ingenieria-robotica-mit-b2207128.html

[9] J. E. Riva. “Introduccion”. Documentation-ROS Wiki. Accedido el 16/02/2025. [Onine].Disponible: https://wiki.ros.org/es/ROS/Introduccion

[10] Allwinner Technology. “R40 Datasheet”. Accedido el 16/02/2025. [Online]. Disponible: https://linux-sunxi.org/images/9/98/Allwinner_R40_Datasheet_V0.1.pdf

Fine-tuning y adaptación de Modelos de Lenguaje abiertos en Infraestructura HPC para aplicaciones de dominio específico

Juan M. Fernandez[1], David Petrocelli[1], Rosana Matuk[1], Daniel Lanson[1], Eduardo Zamudio[2], Leticia Cagnina[3], Verónica Gil Costa[3], Marcelo Errecalde[3]

{jmfernandez, dpetrocelli, rmatuk, dlanson}@unlu.edu.ar,

eduardozamudio@fceqyn.unam.edu.ar, {lcagnina, gvcosta, merreca}@email.unsl.edu.ar

[1]LICDIA, Universidad Nacional de Luján

[2]IIDII, Universidad Nacional de Misiones

[3]LIDIC, Universidad Nacional de San Luis

Resumen

El lenguaje es una característica distintiva de la humanidad, y la conversación es su ámbito fundamental. La interacción conversacional entre máquinas y humanos ha sido una meta en la informática desde hace más de 70 años. Estas capacidades se investigan en el marco de la Inteligencia Artificial Generativa (IAG), que permite a las máquinas generar contenido nuevo y coherente en función de datos de entrenamiento.

Las redes neuronales generativas adversarias y la arquitectura de *transformers*, que revolucionaron el campo, se destacan en la creación de contenido coherente y relevante. Modelos como GPT-4, LLaMA y DeepSeek-V3 han establecido nuevos estándares, demostrando la capacidad de los modelos de lenguaje a gran escala (LLM) para adaptarse a diversas tareas. Los LLM, al ser pre-entrenados con grandes volúmenes de datos, adquieren habilidades generales que pueden ajustarse a nuevas tareas específicas mediante técnicas de *fine-tuning*.

Entre las técnicas de *fine-tuning* más relevantes, el método clásico ajusta todos los parámetros del modelo, siendo efectivo pero computacionalmente costoso. Las estrategias basadas en *adapter tuning*, *compacter*, *BitFit* así como *LoRA* y sus variantes ofrecen alternativas eficientes y modulares. A su vez, enfoques recientes como *Mixture of Experts (MoE)* se constituyen como alternativas para mejorar la eficiencia, activando solo una parte del modelo en cada inferencia.

En este trabajo se describen las acciones emprendidas relacionadas con el *fine-tuning* y adaptación de modelos de lenguaje abiertos en el marco de la presentación a la categoría "Proyectos de Cálculo Intensivo (PCI)" de la Iniciativa de Proyectos Acelerados de Cálculo (IPAC), por medio de la utilización de la **supercomputadora Clementina XXI**, a partir de la colaboración interinstitucional entre el LICDIA-UNLu, el LIDIC-UNSL y el IIDII-UNaM.

Contexto

Esta presentación se encuentra en el marco de la presentación realizada para la categoría "Proyectos de Cálculo Intensivo (PCI)" de la Iniciativa de Proyectos Acelerados de Cálculo (IPAC) para la utilización de la supercomputadora denominada Clementina XXI, con la colaboración interinstitucional del Laboratorio de Investigación en Ciencia de Datos & Inteligencia Artificial (LICDIA) de la Universidad Nacional de Luján, el Laboratorio de Investigación y Desarrollo en Inteligencia Computacional (LIDIC) de la Universidad Nacional de San Luis, y el Instituto de Investigación, Desarrollo e Innovación en Informática (IIDII) de la Universidad Nacional de Misiones.

Introducción

El lenguaje es una de las características más distintivas de la humanidad y la conversación es su principal forma de interacción [17]. A lo largo de más de 70 años, la posibilidad de que una máquina interactúe mediante el lenguaje natural ha sido uno de los grandes desafíos de la informática [28]. Desde los primeros sistemas conversacionales como ELIZA [30] hasta los asistentes modernos como Siri, Alexa y ChatGPT, la inteligencia artificial generativa ha evolucionado significativamente, permitiendo que las máquinas no solo procesen, sino también generen contenido en distintos formatos, como texto, imágenes, sonido y video [9]. Los modelos de lenguaje de gran escala (LLMs) han revolucionado la Inteligencia Artificial Generativa, facilitando la generación de texto altamente coherente y contextualizado. Basados en redes neuronales profundas, estos modelos han pasado por varias etapas de desarrollo, desde las redes neuronales recurrentes (LSTM) [13] y las incrustaciones de palabras (*word embeddings*) [22] hasta las más recientes variaciones de la arquitectura de *transformers* [29], cuyas bases han permitido un aprendi-

zaje eficiente y escalable de dependencias lingüísticas a gran escala.

Sin embargo, los LLMs requieren no solo una gran capacidad computacional, sino también hardware especializado y grandes volúmenes de datos, tanto para su entrenamiento como para la inferencia de las respuestas [18].

Antecedentes

Entre los hitos recientes de mayor importancia en el desarrollo de modelos de procesamiento del lenguaje natural (NLP, por su acrónimo en inglés), destacan BERT [7], un modelo basado en *encoders* que optimiza la representación semántica del texto, y los modelos generativos GPT-4 [1], LLaMA [26] y DeepSeek-V3 [21], que han establecido nuevos estándares en la generación de texto y en tareas de NLP. Preentrenados con grandes volúmenes de datos, estos modelos pueden ajustarse mediante técnicas de *fine-tuning* para adaptarse a tareas específicas con menor cantidad de datos [35].

El ajuste fino de modelos de lenguaje ha evolucionado desde métodos tradicionales hasta enfoques más eficientes en términos computacionales [33].

Por un lado, el *fine-tuning* clásico o convencional implica ajustar todos los parámetros del modelo preentrenado a la tarea específica. Este método es efectivo para adaptar el modelo a nuevas tareas, pero es computacionalmente costoso y requiere grandes cantidades de datos específicos de esa tarea para evitar un sobreajuste [15]. Por su parte, las estrategias basadas en *adapter tuning* introducen pequeñas capas de *"adapters"* (o adaptadores) [14] entre las capas del modelo preentrenado. Estos *adapters* se entrenan para cada nueva situación, mientras que los parámetros originales del modelo se mantienen fijos, lo que permite una adaptación eficiente y modular de los modelos pre-entrenados. A su vez, han aparecido estrategias híbridas como Compacter [19], que combina la idea de a*dapter tuning* con métodos de *low rank* para crear una técnica que es eficiente y efectiva utilizando una descomposición de baja clasificación para reducir el número de parámetros en los *adapters*, mejorando aún más la eficiencia del *fine-tuning*. Por otro lado, existen estrategias como *BitFit (Bias-Only Fine-Tuning)*, que consisten en técnicas minimalistas que ajustan únicamente los parámetros de sesgo (bias) del modelo preentrenado [33]. Aunque estos representan una fracción muy pequeña del total de parámetros, se ha demostrado que pueden capturar información suficiente para adaptar el modelo a nuevas tareas. En el último tiempo, *LoRA (Low-Rank Adaptation)* [16] y algunas de sus variantes como AdaLoRA [34], VB-LoRA [20] o SuperLoRA [4] han ganado popularidad como métodos para el *fine-tuning* de modelos de lenguaje de gran escala (LLM). Estas técnicas han demostrado ser eficientes para ajustar modelos pre-entrenados a tareas específicas sin incurrir en el alto costo computacional asociado con el ajuste de todos los parámetros del modelo introduciendo capas de baja clasificación (*low-rank*) en la arquitectura del modelo, entrenadas para capturar la variabilidad específica de la tarea, mientras que la mayor parte del modelo original permanece congelada.

Además, enfoques como *Mixture of Experts (MoE)* han emergido como alternativas para mejorar la eficiencia, activando solo una parte del modelo en cada inferencia [31].

Sin embargo, los LLMs requieren una gran capacidad computacional tanto para su entrenamiento como para la inferencia [18]. Si bien la computación en la nube ha sido utilizada ampliamente para escalar estos modelos, la computación de alto desempeño (HPC) se presenta como una solución óptima para la optimización del ajuste fino y el entrenamiento distribuido en grandes volúmenes de datos [23]. Herramientas como *DeepSpeed* y *Fully Sharded Data Parallel (FSDP)* permiten distribuir eficientemente el entrenamiento en múltiples GPUs, reduciendo costos y tiempos de cómputo [24]. No obstante, dado que Clementina XXI no cuenta con soporte para CUDA, este proyecto enfrentará el desafío adicional de adaptar estas estrategias a la infraestructura disponible, explorando alternativas en Intel Max Series GPUs mediante oneAPI y SYCL. Esto implica un esfuerzo extra en optimización y configuración para alcanzar niveles de eficiencia similares a los obtenidos en arquitecturas basadas en CUDA.

El aprovechamiento de infraestructuras HPC para el ajuste y optimización de modelos LLMs permitirá mejorar su desempeño en aplicaciones de propósito específico, asegurando soluciones escalables y eficientes para dominios clave como seguridad, información espacial, biología, chatbots especializados, educación y matemática [6].

Líneas de I+D

A continuación se presentan y describen brevemente las principales líneas de I+D en las cuales se trabaja actualmente en el marco de este proyecto.

a. Evaluación y optimización de estrategias de *fine-tuning*

Como se planteó anteriormente, existen diferentes estrategias de adaptación y *fine-tuning* para los modelos de lenguaje de escala, desde el fine-tuning clásico, que ajusta todos los parámetros del modelo para adaptarlo a una tarea específica, pasando por el *adapter tuning* y *compacter* que introduce capas adicionales de parámetros entrenables sin modificar el modelo base, el *BitFit* que modifica únicamente los

parámetros de sesgo del modelo, las estrategias basadas en *LoRA* que permiten una adaptación eficiente sin reentrenar todos los parámetros o las estrategias basadas en MoE que permiten activar solo una parte del modelo en cada inferencia.

En el marco del proyecto, gracias a la capacidad de cómputo mediante GPU de Clementina XXI, se evaluarán y optimizarán estrategias avanzadas de *fine-tuning*, incluyendo LoRA, QLoRA, Adapter Tuning y técnicas emergentes como MoE y Soft Prompting, aprovechando la infraestructura HPC para permitir la adaptación de modelos de lenguaje a tareas específicas y mejorar la eficiencia computacional en términos de tiempos de entrenamiento y uso de recursos.

b. Entrenamiento de modelos de propósito específico

Los modelos de lenguaje de gran escala (LLMs) han demostrado capacidades excepcionales en tareas generales de procesamiento del lenguaje natural (NLP, por su acrónimo en inglés). Sin embargo, su desempeño en dominios específicos aún enfrenta desafíos debido a la escasez de datos especializados y la posible interferencia de conocimientos adquiridos durante el pre-entrenamiento, lo que puede reducir la precisión en contextos altamente especializados [5][8]. Estudios recientes han evidenciado que la adaptación de LLMs a contextos especializados requiere estrategias que mitiguen la interferencia contextual y optimicen la incorporación de terminología específica del dominio [3]. A su vez, existe evidencia sobre los riesgos de la degradación del rendimiento cuando los modelos son entrenados excesivamente con datos generados por otras IAs, fenómeno conocido como *model collapse*, en el cual las distribuciones originales de los datos comienzan a perderse progresivamente con cada nueva iteración de entrenamiento [25][12].

Para abordar estas limitaciones, en el marco de esta Línea de I+D, se exploran estrategias avanzadas de entrenamiento y adaptación de modelos, combinando *LangChain* para la integración de flujos de trabajo específicos y *LlamaIndex* para la indexación y recuperación eficiente de datos de dominios específicos.

Asimismo, y dado que la infraestructura de Clementina XXI se basa en las *Intel Data Center GPU Max Series*, este trabajo explora la utilización de *oneAPI* y *SYCL* como alternativas a los frameworks tradicionales basados en *CUDA*. Estas GPUs, basadas en la arquitectura *Ponte Vecchio*, han sido diseñadas para cargas de trabajo intensivas en HPC e IA, incorporando memoria HBM2e de alto ancho de banda y unidades especializadas XMX (Xe Matrix eXtensions) para la aceleración de operaciones matriciales en modelos de lenguaje de gran escala [2]. En este contexto, la optimización del entrenamiento distribuido y la gestión eficiente de memoria juegan un papel clave en la escalabilidad de LLMs en entornos de computación de alto rendimiento [32].

c. Desarrollo de soluciones basadas en LLM

El despliegue de modelos de lenguaje de gran escala (LLMs) en entornos de producción plantea desafíos tanto a nivel computacional como en términos de alineación con tareas específicas. A medida que estos modelos incrementan su número de parámetros, optimizar su ejecución distribuida y la gestión eficiente de recursos computacionales se vuelve esencial para garantizar su escalabilidad [27][11].

Para abordar estos desafíos, se han desarrollado enfoques de paralelismo distribuido que optimizan la eficiencia en entornos de ejecución a gran escala. Métodos como *Fully Sharded Data Parallel (FSDP)* han demostrado ser efectivos para distribuir el entrenamiento de modelos [32]. Estas optimizaciones resultan fundamentales para el despliegue eficiente de LLMs en escenarios de alto rendimiento, asegurando su aplicabilidad en múltiples sectores.

En el marco de esta Línea de I+D, se exploran herramientas avanzadas para la integración de LLMs en aplicaciones prácticas. La librería Transformers de Hugging Face permite gestionar y optimizar modelos preentrenados con soporte para técnicas de inferencia eficiente, como la cuantización [10], que reduce el costo computacional sin comprometer significativamente la calidad de las predicciones. Asimismo, se utilizan librerías como *LangChain*, que permiten la construcción modular de agentes conversacionales y sistemas de recomendación personalizados, facilitando la integración de LLMs en flujos de trabajo estructurados y adaptados a diferentes dominios.

Objetivos

El objetivo general del Proyecto es explorar y desarrollar estrategias eficientes de fine-tuning y adaptación de modelos de lenguaje abiertos en infraestructura HPC, optimizando su aplicación en diversos dominios, con especial énfasis en la implementación de soluciones reales.

A partir del objetivo general, se definen los siguientes objetivos específicos:

- Evaluar y optimizar estrategias avanzadas de fine-tuning, incluyendo LoRA, QLoRA, Adapter Tuning y técnicas emergentes como MoE y Soft Prompting, aprovechando la infraestructura HPC para permitir la adaptación de modelos de lenguaje a tareas específicas y mejorar la eficiencia computacional en términos de tiempos de entrenamiento y uso de recursos.

- Establecer criterios cuantificables para comparar modelos ajustados en español y medir su desempeño en tareas específicas.
- Implementar modelos de lenguaje adaptados en casos de uso concretos, priorizando aplicaciones en seguridad, información ambiental, espacial y biológica así como chatbots sobre dominios específicos, educación y matemática, sin excluir otras áreas donde la optimización y adaptación de estos modelos pueda resultar beneficiosa.
- Generar repositorios abiertos, publicar artículos científicos y compartir herramientas prácticas para facilitar la adopción y replicabilidad de los resultados.

Recursos Humanos

Se espera que este proyecto contribuya a seguir consolidando un grupo de investigación interinstitucional entre el Laboratorio de Investigación en Ciencia de Datos & Inteligencia Artificial (LICDIA, UNLu) el Laboratorio de Investigación y Desarrollo en Inteligencia Computacional (LIDIC, UNSL) y el Instituto de Investigación, Desarrollo e Innovación en Informática (IIDII) de la Universidad Nacional de Misiones en el área de Inteligencia Artificial aplicada.

Al mismo tiempo, se espera que este proyecto brinde un ámbito adecuado para la formación de recursos humanos que incorporen saberes y competencias provenientes de la participación en actividades de investigación.

Concretamente, se esperan incluir al menos tres auxiliares docentes como integrantes para el desarrollo del Proyecto, al mismo tiempo que se espera incorporar al menos un pasante, en el marco del Programa de Pasantías Internas Rentadas (PIR) (Resolución RESHCS-LUJ: 0000116-19), de la UNLu a partir de la radicación del proyecto.

Referencias

[1] Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774* (2023).

[2] Applencourt, T., Sadawarte, A., Muralidharan, S., Bertoni, C., Kwack, J., Luo, Y., Rangel, E., Tramm, J., Ghadar, Y., Tamerus, A., et al. Ponte vecchio across the atlantic: Single-node benchmarking of two intel gpu systems. In *SC24-W: Workshops of the International Conference for High Performance Computing, Networking, Storage and Analysis* (2024), IEEE, pp. 1430–1442.

[3] Chalkidis, I., Jana, A., Hartung, D., Bommarito, M., Androutsopoulos, I., Katz, D. M., and Aletras, N. Lexglue: A benchmark dataset for legal language understanding in english. *arXiv preprint arXiv:2110.00976* (2021).

[4] Chen, X., Liu, J., Wang, Y., Wang, P. P., Brand, M., Wang, G., and Koike-Akino, T. Superlora: Parameter-efficient unified adaptation of multi-layer attention modules. *arXiv preprint arXiv:2403.11887* (2024).

[5] Coleman, E. N., Hurtado, J., and Lomonaco, V. In-context interference in chat-based large language models. In *European Robotics Forum* (2024), Springer, pp. 114–119.

[6] Cruz, L., Gutierrez, X. F., and Martínez-Fernández, S. Innovating for tomorrow: The convergence of se and green ai. *arXiv preprint arXiv:2406.18142* (2024).

[7] Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)* (2019), pp. 4171–4186.

[8] Eitan, D., Pirchi, M., Glazer, N., Meital, S., Ayach, G., Krendel, G., Shamsian, A., Navon, A., Hetz, G., and Keshet, J. Combining language models for specialized domains: A colorful approach. *arXiv preprint arXiv:2310.19708* (2023).

[9] Feuerriegel, S., Hartmann, J., Janiesch, C., and Zschech, P. Generative ai. *Business & Information Systems Engineering 66*, 1 (2024), 111–126.

[10] Frantar, E., and Alistarh, D. Optimal brain compression: A framework for accurate post-training quantization and pruning. *Advances in Neural Information Processing Systems 35* (2022), 4475–4488.

[11] Frantar, E., Ashkboos, S., Hoefler, T., and Alistarh, D. Gptq: Accurate post-training quantization for generative pre-trained transformers. *arXiv preprint arXiv:2210.17323* (2022).

[12] Gerstgrasser, M., Schaeffer, R., Dey, A., Rafailov, R., Sleight, H., Hughes, J., Korbak, T., Agrawal, R., Pai, D., Gromov, A., et al. Is model collapse inevitable? breaking the curse of recursion by accumulating real and synthetic data. *arXiv preprint arXiv:2404.01413* (2024).

[13] Hochreiter, S., and Schmidhuber, J. Long short-term memory. *Neural computation 9*, 8 (1997), 1735–1780.

[14] Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., De Laroussilhe, Q., Gesmundo, A., Attariyan, M., and Gelly, S. Parameter-efficient transfer learning for nlp. In *International conference on machine learning* (2019), PMLR, pp. 2790–2799.

[15] Howard, J., and Ruder, S. Universal language model fine-tuning for text classification. *arXiv preprint arXiv:1801.06146* (2018).

[16] Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., et al. Lora: Low-rank adaptation of large language models. *ICLR 1*, 2 (2022), 3.

[17] Jurafsky, D., and Martin, J. H. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models*, 3rd ed. 2025. Online manuscript released January 12, 2025.

[18] Kachris, C. A survey on hardware accelerators for large language models. *Applied Sciences 15*, 2 (2025), 586.

[19] Karimi Mahabadi, R., Henderson, J., and Ruder, S. Compacter: Efficient low-rank hypercomplex adapter layers. *Advances in Neural Information Processing Systems 34* (2021), 1022–1035.

[20] Li, Y., Han, S., and Ji, S. Vb-lora: extreme parameter efficient fine-tuning with vector banks. *arXiv preprint arXiv:2405.15179* (2024).

[21] Liu, A., Feng, B., Xue, B., Wang, B., Wu, B., Lu, C., Zhao, C., Deng, C., Zhang, C., Ruan, C., et al. Deepseek-v3 technical report. *arXiv preprint arXiv:2412.19437* (2024).

[22] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems 26* (2013).

[23] Patil, P., and Pandhare, S. Y. Literature survey on cloud computing and its benefits. *International Journal of Research Publication and Reviews 3*, 11 (November 2022), 2901–2904.

[24] Sevilla, J., Heim, L., Hobbhahn, M., Besiroglu, T., Ho, A., and Villalobos, P. Estimating training compute of deep learning models. *Tech. Rep.* (2022).

[25] Shumailov, I., Shumaylov, Z., Zhao, Y., Gal, Y., Papernot, N., and Anderson, R. The curse of recursion: Training on generated data makes models forget. *arXiv preprint arXiv:2305.17493* (2023).

[26] Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., et al. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971* (2023).

[27] Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S., et al. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288* (2023).

[28] Turing, A. M. Computing machinery and intelligence. *Mind LIX*, 236 (1950), 433–460.

[29] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. Attention is all you need. *Advances in neural information processing systems 30* (2017).

[30] Weizenbaum, J. Eliza—a computer program for the study of natural language communication between man and machine. *Communications of the ACM 9*, 1 (1966), 36–45.

[31] Yang, S., Ali, M. A., Wang, C.-L., Hu, L., and Wang, D. Moral: Moe augmented lora for llms' lifelong learning. *arXiv preprint arXiv:2402.11260* (2024).

[32] Yuan, K., Bauinger, C., Zhang, X., Baehr, P., Kirchhart, M., Dabert, D., Tousnakhoff, A., Boudier, P., and Paulitsch, M. Fully-fused multi-layer perceptrons on intel data center gpus. *arXiv preprint arXiv:2403.17607* (2024).

[33] Zaken, E. B., Ravfogel, S., and Goldberg, Y. Bitfit: Simple parameter-efficient fine-tuning for transformer-based masked language-models. *arXiv preprint arXiv:2106.10199* (2021).

[34] Zhang, Q., Chen, M., Bukharin, A., Karampatziakis, N., He, P., Cheng, Y., Chen, W., and Zhao, T. Adalora: Adaptive budget allocation for parameter-efficient fine-tuning. *arXiv preprint arXiv:2303.10512* (2023).

[35] Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., et al. A survey of large language models. *arXiv preprint arXiv:2303.18223 1*, 2 (2023).

Integración de Grafos de Conocimiento y Modelos RAG para la Gestión Inteligente de Datos Oceanográficos

Marcos Zárate[1,2,3], Pablo Fillottrani[4], Carlos Buckle[1,2], Dario Ceballos[3,5], Gustavo Nuñez[3], Macarena Repetto[1,2], and Fabián Vidal[1]

[1]Facultad de Ingeniería, Universidad Nacional de la Patagonia San Juan Bosco (FI-UNPSJB)
[2]Laboratorio de Investigación en Informática de la UNPSJB (LINVI-UNPSJB)
[3]Centro para el Estudio de Sistemas Marinos, Centro Nacional Patagónico,
Consejo Nacional de Investigaciones Científicas y Técnicas (CESIMAR-CENPAT-CONICET)
[4]Laboratorio de Investigación y Desarrollo en Ingeniería de Software y Sistemas de Información (LISSI-DCIC-UNS)
[5]Facultad de Tecnología Informática, Universidad Abierta Interamericana (FTI-UAI)
zarate@cenpat-conicet.gob.ar, prf@cs.uns.edu.ar, cbuckle@unpata.edu.ar,
dceballos@cenpat-conicet.gob.ar, gnunez@cenpat-conicet.gob.ar, macarepetto@gmail.com,
vidalfabian013@gmail.com

Resumen

La Inteligencia Artificial (IA) está transformando la ciencia al acelerar el análisis de datos, facilitar descubrimientos, mejorar diagnósticos y pronósticos, automatizar tareas y promover la colaboración. Esta revolución científica abre oportunidades para la investigación y el descubrimiento en diversas disciplinas como las ciencias del mar. Específicamente esta investigación propone integrar el grafo de conocimiento OceanGraph, previamente desarrollado por las instituciones participantes con modelos de lenguaje grandes (LLMs) y técnicas de recuperación asistida por grafos (GraphRAG) para mejorar la recuperación de datos y generar conocimientos más precisos y contextualizados en el campo de la oceanografía. Con esto, se espera avanzar tanto en la ciencia oceanográfica como en la aplicación de tecnologías de inteligencia artificial para la gestión y análisis de grandes conjuntos de datos marinos.

Palabras clave: Grafos de conocimiento, Modelos RAG, Oceanografía, IA, Datos enlazados.

Contexto

La oceanografía ha generado una cantidad masiva de datos en las últimas décadas. Sin embargo, estos datos suelen permanecer fragmentados en diferentes repositorios, dificultando su acceso y utilización. OceanGraph surge como una iniciativa para la estructuración y vinculación de estos datos mediante grafos de conocimiento, permitiendo la recuperación eficiente de información y facilitando su integración con modelos RAG para mejorar el análisis y la toma de decisiones en investigaciones científicas. Esta propuesta avanzará sobre resultados y líneas de investigación del proyecto "Extracción y Explotación de Conocimiento para la gestión en línea de datos en ciencias del mar", en el cual se generó un grafo de conocimiento (KG) denominado OceanGraph, que contiene información oceanográfica de diferentes fuentes de datos.

El proyecto tiene una concepción interdisci-

plinaria y por ello la unidad ejecutora incluye investigadores de las ciencias de la computación y de las ciencias del mar, pertenecientes a CESIMAR-CONICET, LINVI-UNPSJB, FTI-UAI y LISSI-DCIC-UNS. El proyecto fue avalado y financiado por la UNPSJB para ser ejecutado entre 2024 y 2026.

1. Introducción

El aumento de las temperaturas oceánicas afecta severamente a las especies marinas y a los ecosistemas. Durante agosto de 2023 se observaron las temperaturas superficiales más altas jamás registradas [1], lo cual es alarmante por muchas razones (aunque quizás no sea sorprendente considerando que el océano absorbe el 90 % del exceso de calor del planeta) y es el mayor depósito de carbono del mundo. Pero entre más caliente esté el aire, más se calentarán inevitablemente los océanos. Fenómenos como el blanqueamiento de corales [2] y la pérdida de hábitats de reproducción para peces y mamíferos marinos [3] son solo algunos de los impactos observados. Frente a estos problemas globales, los investigadores de todo el mundo han generado enormes cantidades de datos oceanográficos en los últimos años [4]. Sin embargo, la mayoría de estos datos permanecen almacenados de manera que no pueden ser fácilmente compartidos o integrados con otros sistemas o bases de datos [5], lo que dificulta su accesibilidad y análisis. OceanGraph [6, 7] representa un esfuerzo por superar estas barreras, proporcionando un espacio de conocimiento compartido donde entidades como publicaciones, personas, lugares, especímenes, variables ambientales e instituciones están interconectadas. Utilizando los principios de los datos enlazados, OceanGraph permite a los usuarios consultar y razonar sobre algunos de los mayores repositorios de datos oceanográficos y de biodiversidad marina. Sin embargo existen limitantes debido a que los grafos de conocimiento, aunque potentes, presentan varias limitaciones [8, 9]. La escalabilidad y la complejidad computacional son desafíos significativos [10]. A medida que se añaden más entidades y relaciones, el tamaño del grafo crece considerablemente, lo que incrementa la carga computacional y puede afectar el rendimiento durante la consulta y actualización de datos. Además, mantener estos grafos actualizados y precisos es una tarea ardua, especialmente en dominios donde la información cambia rápidamente. La integración de nuevas fuentes de datos y la eliminación de información obsoleta requieren procesos robustos y automáticos, que son difíciles de implementar [11]. Otra limitación importante es la ambigüedad y la falta de precisión. Cuando se integran datos de múltiples fuentes con diferentes niveles de calidad y formatos, los grafos de conocimiento pueden enfrentar problemas de inconsistencia [12]. Resolver estas inconsistencias y asegurar que las relaciones entre entidades sean precisas y relevantes es una tarea compleja. Asimismo, los grafos de conocimiento están limitados por la semántica predefinida y las relaciones explícitas entre entidades. Esto restringe su capacidad para inferir nuevas relaciones o entender contextos complejos que no están explícitamente representados en el grafo. Para superar estas limitaciones, los sistemas RAG [13] y los LLMs [14] ofrecen soluciones efectivas. Los LLMs como GPT [15], Llama [16], DeepSeek [17], son avanzados en la comprensión y generación de lenguaje natural en diversos contextos, lo cual les permite generar respuestas contextualmente adecuadas y relevantes. Estos modelos son capaces de inferir información y descubrir relaciones implícitas a partir de grandes volúmenes de datos no estructurados, lo cual amplía su aplicabilidad en la recuperación y generación de información. Sin embargo, los LLMs tienen limitaciones, especialmente en términos de control sobre la precisión y consistencia de las respuestas, pues dependen de patrones en el texto no estructurado y carecen de la estructura explícita que ofrecen los grafos de conocimiento. Además, su capacidad para manejar información actualizada en tiempo real es limitada y puede depender de procesos de retraining costosos y esporádicos. Para mitigar estas limitaciones, surgen los sistemas RAG, estos sistemas combinan la recuperación de información con la generación de

texto: primero recuperan información relevante de una fuente estructurada o semiestructurada, como un grafo de conocimiento, y luego generan respuestas basadas en el contexto. Al incorporar RAG, es posible que los LLMs generen respuestas más precisas y contextuales al complementarse con fuentes estructuradas. En este enfoque, los modelos de lenguaje procesan la información recuperada y, mediante sus capacidades generativas, entregan respuestas que son tanto informativas como contextualmente adecuadas. En términos de escalabilidad y actualización, los LLMs manejan grandes volúmenes de datos y generan respuestas rápidamente, lo cual ayuda a superar algunos problemas de escalabilidad de los grafos de conocimiento. Adicionalmente, los sistemas RAG permiten la actualización dinámica de la información recuperada, asegurando que las respuestas generadas estén alineadas con la información actualizada, algo difícil de lograr con LLMs por sí solos. La integración de OceanGraph en este tipo de sistema sería altamente beneficiosa, ya que proporcionará una estructura sólida de información y relaciones explícitas que los LLMs podrían usar como base de referencia. Esta combinación permitiría una mayor precisión y coherencia en las respuestas generadas, ya que OceanGraph complementaría las capacidades inferenciales y generativas de los LLMs con una estructura confiable y constantemente actualizada. Así, la sinergia entre OceanGraph y los sistemas RAG apoyados por LLMs ofrecería respuestas más robustas, relevantes y adecuadas a contextos específicos, superando muchas de las limitaciones inherentes tanto de los grafos de conocimiento tradicionales como de los LLMs.

2. Lineas de Investigación y desarrollo

El presente proyecto abarca diversas líneas de investigación centradas en la integración de grafos de conocimiento con modelos de inteligencia artificial para la gestión de datos oceanográficos. En primer lugar, se trabaja en el desarrollo y optimización de OceanGraph como un grafo de conocimiento interoperable, utilizando principios de datos enlazados para mejorar la integración y consulta de información. Además, se investiga la aplicación de modelos de Generación Aumentada por Recuperación (RAG) para optimizar la contextualización y precisión de respuestas en el análisis de datos oceanográficos, incorporando técnicas avanzadas de prompting y explainability.

Otra línea clave es el análisis y predicción de datos oceanográficos mediante inteligencia artificial, explorando el uso de modelos de aprendizaje automático para detectar patrones en series temporales y generar modelos híbridos que combinen enfoques simbólicos y subsimbólicos. También se abordan desafíos en infraestructura computacional y escalabilidad, desarrollando arquitecturas eficientes para la consulta y almacenamiento de grandes volúmenes de datos, optimizando la recuperación semántica y facilitando la integración de OceanGraph con plataformas en la nube.

Finalmente, el proyecto busca aplicaciones científicas concretas en estudios de biodiversidad marina y conservación de especies, promoviendo la generación de herramientas de visualización y análisis que faciliten la toma de decisiones en investigaciones oceanográficas. Se fomenta, además, la colaboración interdisciplinaria entre especialistas en ciencias de la computación y oceanografía, impulsando nuevas sinergias y conocimientos en el ámbito de la gestión inteligente de datos científicos.

3. Resultados esperados

- Evaluar la efectividad de OceanGraph para mejorar la precisión y relevancia contextual de los modelos RAG.
- Aumentar la capacidad de prompting y explainability mediante el contexto que otorgan los grafos de conocimiento.
- Probar y validar el modelo OceanGraph en diversos escenarios de aplicación, como la predicción de fenómenos climáticos y el monitoreo de la biodiversidad marina.

- Desarrollar herramientas metodológicas replicables en otras áreas científicas y tecnológicas.

4. Formación de Recursos Humanos

El presente proyecto apunta a sostener y fortalecer las capacidades de los grupos de investigación LINVI-UNPSJB, CESIMAR-CONICET y LISSI-DCIC-UNS que trabajan en forma colaborativa desde hace 9 años en los campos de modelado conceptual, desarrollo de ontologías de dominio y, más recientemente, en temáticas de Knowledge Graph, RAG y LLMs. Como la mayoría de los grupos que investigan en Ciencias de Computación/Informática y que pertenecen a unidades ejecutoras de CONICET, se trata de grupos pequeños que no han podido crecer como se esperaba, más bien han reducido sus tamaños, en los últimos años. En efecto, los recursos humanos más jóvenes que los integraban han sido captados por la industria de software y servicios informáticos. Por otra parte, debe recalcarse que en el país prácticamente no existen grupos de investigación que investiguen en las temáticas de este proyecto, que son consideradas prioritarias en el plan nacional de ciencia, tecnología e innovación 2030, elaborado en el ámbito del MINCyT. En consecuencia, resulta imperioso sostener y fortalecer las capacidades actuales en términos de recursos humanos para no perder los esfuerzos realizados en las últimas años y fortalecer las líneas de investigación vigentes. Asimismo, el becario doctoral Gustavo Nuñez inicia el desarrollo de sus Tesis Doctoral titulada "Inteligencia artificial como herramienta para la explotación de datos oceanográficos", mientras que el becario doctoral Darío Ceballos avanzará con su tesis titulada "Analíticos Visuales para Datos Enlazados en Ciencias del Mar" en el marco de este proyecto.

Ademas participan 2 alumnos de la Licenciatura en Informática de la UNPSJB, Macarena Repetto con beca ECV-CIN cuyo plan de trabajo se titula "Exploración de Indicadores Ambientales y Poblacionales de Especies Marinas mediante Inteligencia Artificial y Visualización de Datos" y Fabián Vidal quien colabora con el desarrollo de aplicaciones web.

Además, se espera que la temática del mismo resulte de interés para el desarrollo de becas de investigación para alumnos de grado que van a ser ofrecidas (en la UNPSJB que poseen programas que las financian, además de las becas EVC-CIN), así como proyectos de fin de carrera.

Referencias

[1] Boyin Huang, Xungang Yin, James A Carton, Ligang Chen, Garrett Graham, Patrick Hogan, Thomas Smith, and Huai-Min Zhang. Record high sea surface temperatures in 2023. *Geophysical Research Letters*, 51(14):e2024GL108369, 2024.

[2] AE Douglas. Coral bleaching—-how and why? *Marine pollution bulletin*, 46(4):385–392, 2003.

[3] Peter GH Evans and Arne Bjørge. Impacts of climate change on marine mammals. *MCCIP Science Review*, 2013:134–148, 2013.

[4] Yingjian Liu, Meng Qiu, Chao Liu, and Zhongwen Guo. Big data in ocean observation: Opportunities and challenges. In *Big Data Computing and Communications: Second International Conference, BigCom 2016, Shenyang, China, July 29-31, 2016. Proceedings 2*, pages 212–222. Springer, 2016.

[5] Adèle Révelard, Joaquín Tintoré, Jacques Verron, Pierre Bahurel, John A Barth, Mathieu Belbéoch, Jérôme Benveniste, Pascal Bonnefond, Eric P Chassignet, Sophie Cravatte, et al. Ocean integration: The needs and challenges of effective coordination within the ocean observing system. *Frontiers in Marine Science*, 8:737671, 2022.

[6] Marcos Zárate, Pablo Rosales, Germán Braun, Mirtha Lewis, Pablo Rubén Fillottrani, and Claudio Delrieux. Oceangraph: some initial steps toward a oceanographic knowledge graph. In *Knowledge Graphs and Semantic Web: First Iberoamerican Conference, KGSWC 2019, Villa Clara, Cuba, June 23-30, 2019, Proceedings 1*, pages 33–40. Springer, 2019.

[7] Marcos Zárate, Carlos Buckle, Renato Mazzanti, Mirtha Lewis, Pablo Fillottrani, and Claudio Delrieux. Harmonizing big data with a knowledge graph: Oceangraph kg uses case. In *Conference on Cloud Computing, Big Data & Emerging Topics*, pages 81–92. Springer, 2020.

[8] Xiangyu Wang, Lyuzhou Chen, Taiyu Ban, Muhammad Usman, Yifeng Guan, Shikang Liu, Tianhao Wu, and Huanhuan Chen. Knowledge graph quality control: A survey. *Fundamental Research*, 1(5):607–626, 2021.

[9] Ruojing Zhang, Marta Indulska, and Shazia Sadiq. Discovering data quality problems: the case of repurposed data. *Business & Information Systems Engineering*, 61:575–593, 2019.

[10] Gleb Gawriljuk, Andreas Harth, Craig A Knoblock, and Pedro Szekely. A scalable approach to incrementally building knowledge graphs. In *Research and Advanced Technology for Digital Libraries: 20th International Conference on Theory and Practice of Digital Libraries, TPDL 2016, Hannover, Germany, September 5–9, 2016, Proceedings 20*, pages 188–199. Springer, 2016.

[11] Enrique Iglesias, Samaneh Jozashoori, and Maria-Esther Vidal. Scaling up knowledge graph creation to large and heterogeneous data sources. *Journal of Web Semantics*, 75:100755, 2023.

[12] P Heyvaert, B De Meester, A Dimou, and R Verborgh. Rule-driven inconsistency resolution for knowledge graph generation rules. semant. web. 10,(2019).

[13] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33:9459–9474, 2020.

[14] Arun James Thirunavukarasu, Darren Shu Jeng Ting, Kabilan Elangovan, Laura Gutierrez, Ting Fang Tan, and Daniel Shu Wei Ting. Large language models in medicine. *Nature medicine*, 29(8):1930–1940, 2023.

[15] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

[16] Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, et al. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*, 2024.

[17] Aixin Liu, Bei Feng, Bing Xue, Bingxuan Wang, Bochao Wu, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan, et al. Deepseek-v3 technical report. *arXiv preprint arXiv:2412.19437*, 2024.

Inteligencia Artificial para la Predicción de Riesgo Cardiovascular en pacientes de OSEP

Rotella Carina[2], Tagarelli Sandra[1,2], Gómez Portillo Juan Manuel, Ontiveros Patricia[2], Monetti Julio, Caymes-Scutari Paola[3,4], Bianchini Germán[3]

[1]Laboratorio de Analítica de Datos, Departamento de Ingeniería en Sistemas de Información.
Facultad Regional Mendoza/Universidad Tecnológica Nacional
Rodríguez 273 (M5502AJE) Mendoza
[2]Laboratorio de Gobierno Electrónico, Departamento de Ingeniería en Sistemas de Información.
Facultad Regional Mendoza/Universidad Tecnológica Nacional
Rodríguez 273 (M5502AJE) Mendoza
[3]Laboratorio de Investigación en Cómputo Paralelo/Distribuido
Departamento de Ingeniería en Sistemas de Información
Facultad Regional Mendoza/Universidad Tecnológica Nacional
Rodríguez 273 (M5502AJE) Mendoza, +54 261 5244579
[4]Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)

crotella@frm.utn.edu.ar, stagarelli@frm.utn.edu.ar, jmgomezportillo@gmail.com, pontiveros@frm.utn.edu.ar, julio.monetti@gmail.com, pcaymesscutari@frm.utn.edu.ar, gbianchini@frm.utn.edu.ar

RESUMEN

Las enfermedades cardiovasculares son la principal causa de morbilidad y mortalidad en Argentina [2]. La Obra Social de Empleados Públicos (OSEP) de Mendoza dispone de un sistema de historia clínica digital (SISAO) con más de 10 años de datos epidemiológicos y clínicos de sus 400.000 afiliados. Este proyecto busca desarrollar un modelo predictivo basado en inteligencia artificial para calcular el riesgo cardiovascular a 10 y 30 años, ajustado a la población local. La implementación de esta herramienta permitirá identificar pacientes de alto riesgo, optimizar la prevención y mejorar la asignación de recursos sanitarios.

Palabras clave: Inteligencia Artificial, Riesgo Cardiovascular, Prevención, Epidemiología, OSEP

CONTEXTO

El proyecto se enmarca en las iniciativas de OSEP para la mejora de la prevención cardiovascular mediante tecnologías avanzadas. Se desarrolla en colaboración con el área de Tecnologías de la Información y Comunicación (TICs) de OSEP y cuenta con el respaldo de su red de efectores de salud.

Tal como relata D'Agostino [4] podemos encontrar en la actualidad múltiples algoritmos de riesgo multivariable para evaluar el riesgo de eventos específicos de enfermedades cardiovasculares ateroscleróticas (ECV), como la enfermedad coronaria, enfermedad cerebrovascular, enfermedad vascular periférica e insuficiencia cardíaca.

La predicción del riesgo cardiovascular ha avanzado significativamente con el desarrollo

de modelos matemáticos y herramientas de inteligencia artificial. Tradicionalmente, modelos como el de Framingham [1] han permitido estimar el riesgo a 10 y 30 años basándose en factores como edad, presión arterial, colesterol, tabaquismo y diabetes. Sin embargo, estos modelos presentan limitaciones en poblaciones diversas y en la integración de datos clínicos más complejos.

El uso de machine learning y deep learning ha permitido mejorar la precisión de estos modelos. Estudios recientes han demostrado que redes neuronales pueden analizar electrocardiogramas para detectar insuficiencia cardíaca con precisión similar a la de un cardiólogo [5] y que algoritmos de IA pueden identificar arritmias a partir de grandes volúmenes de datos clínicos [6].

Además, la inteligencia artificial ha sido aplicada en el análisis de historias clínicas electrónicas para detectar patrones de riesgo cardiovascular con mayor eficiencia [7], y en imágenes médicas, donde el deep learning ha permitido predecir el riesgo cardiovascular a partir de fotografías de retina [8].

Estos avances indican que la combinación de modelos tradicionales con herramientas de IA podría mejorar significativamente la detección temprana y la prevención de enfermedades cardiovasculares, optimizando la asignación de recursos y el tratamiento personalizado en la atención primaria.

1. INTRODUCCIÓN

Las enfermedades cardiovasculares representan un problema de salud pública de alta relevancia. Estudios internacionales, como el de Framingham, han demostrado que factores como edad, hipertensión, diabetes, tabaquismo y obesidad influyen en el riesgo cardiovascular. OSEP, con SISAO [3], ofrece una oportunidad única para la aplicación de modelos de machine learning en la predicción de estos riesgos.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El presente trabajo se enmarca en la rama de Agentes y Sistemas Inteligentes. En primer lugar, como herramienta, se aplicarán modelos de Analítica de Datos y Aprendizaje automático, con el objetivo de segmentar los pacientes en función de características que permitan agruparlos en función del riesgo cardiovascular. Como disciplina, la Analítica de datos puede aplicarse a todas las disciplinas del desempeño humano.

A su vez, también se contemplan otras ramas de la Inteligencia artificial: Un motor de recomendación basado en IA podría sugerir estrategias personalizadas de prevención cardiovascular según el perfil de riesgo de cada paciente. Se explora la posibilidad de aplicar procesamiento de lenguaje natural para interpretar textos en historias clínicas [9].

Finalmente, se prevé la integración a los sistemas de salud existentes.

3. RESULTADOS OBTENIDOS/ESPERADOS

El desarrollo de este proyecto busca aplicar inteligencia artificial y técnicas avanzadas de análisis de datos para mejorar la prevención de enfermedades cardiovasculares en la población afiliada a OSEP. A través del uso de datos históricos de SISAO y modelos de machine learning, se pretende identificar factores de riesgo con mayor precisión, optimizar la detección temprana de pacientes en riesgo y mejorar la asignación de recursos en el sistema de salud [11].

Para poder conformar un set de datos significativo que permita analizar y encontrar variables predictoras de eventos cardiovasculares o muerte de origen cardiovascular en un período de 10 años, se utilizará la siguiente estructura basada en la información disponible en SISAO:

Periodo de análisis:

Fecha de inicio: 01/01/2013
Fecha de fin: 31/12/2023

Variables a Extraer de SISAO

1. **Antecedentes Personales:**
    - Hipertensión arterial
    - Diabetes
    - Tabaquismo
    - Dislipidemia
    - Sedentarismo (falta de actividad física)
2. **Antecedentes Heredofamiliares:**
    - Cardiopatía isquémica prematura
3. **Hábitos:**
    - Consumo de tabaco
4. **Registros en Examen Físico:**
    - Peso
    - Talla
    - Presión arterial sistólica
    - Presión arterial diastólica
5. **Consumo de Medicamentos (según códigos ATC):**
    - **Consumo de al menos 3 recetas** de medicamentos con los siguientes códigos ATC:
        - C02, C03, C07, C08, C09
    - **Consumo de al menos 3 recetas** de medicamentos con el código ATC:
        - A10

Estructura de la Tabla de Eventos Cardiovasculares

La tabla destinada a capturar eventos cardiovasculares deberá contener los siguientes campos:

- **Id temovafi:** Identificador único de afiliado
- **Carácter Externo:** Tipo de efector
- **Institución:** Nombre de la institución de atención
- **Domicilio de atención:** Dirección donde se presta la atención
- **Departamento domicilio de atención:** Departamento correspondiente
- **Número de afiliado:** Número de registro del afiliado
- **Vinculación afiliado:** Tipo de vinculación del afiliado con OSEP
- **DNI afiliado:** Documento Nacional de Identidad
- **Afiliado apellido:** Apellido del afiliado
- **Afiliado nombre:** Nombre del afiliado
- **Fecha de nacimiento:** Fecha de nacimiento del afiliado
- **Sexo:** Género del afiliado
- **Departamento del afiliado:** Departamento de residencia
- **Fecha de inicio:** Fecha de inicio del evento o registro
- **Fecha de fin:** Fecha de finalización del evento o registro
- **Motivo Fin:** Motivo de cierre del registro (por ejemplo, alta médica, traslado)
- **Días de estancia:** Número de días de hospitalización o seguimiento
- **Código diagnóstico OSEP:** Código asignado al diagnóstico según el sistema OSEP
- **Diagnóstico:** Descripción del diagnóstico

- **Práctica:** Procedimientos o prácticas realizadas
- **Motivo de egreso:** Motivo de egreso (por ejemplo, óbito)

Esta estructura permitirá identificar y analizar de forma integral los factores asociados al riesgo cardiovascular en la población afiliada, proporcionando una base robusta para el entrenamiento y validación de los modelos predictivos basados en inteligencia artificial.

Los resultados esperados incluyen la creación de un modelo predictivo de riesgo cardiovascular, su integración en la atención primaria, la implementación de un sistema de alerta temprana y la generación de scores personalizados para la toma de decisiones clínicas. Estos avances permitirán fortalecer las estrategias de prevención y mejorar la calidad de atención de los afiliados, promoviendo un enfoque de salud basado en datos y tecnología.

_Desarrollo de un modelo predictivo de riesgo cardiovascular: se espera la construcción de un modelo predictivo de riesgo cardiovascular basado en datos históricos de SISAO, utilizando técnicas avanzadas de machine learning.

_Implementación de técnicas de machine learning: la aplicación de algoritmos de clasificación y segmentación permitirá identificar patrones de alto riesgo y estratificar a los pacientes según su probabilidad de sufrir eventos cardiovasculares.

_Evaluación y validación del modelo: el modelo se evaluará y validará con cohortes internas de OSEP para asegurar su eficacia y ajustar los parámetros según las características específicas de la población.

_Integración del modelo con el sistema de salud: la herramienta se integrará al sistema de atención primaria de OSEP, permitiendo una aplicación práctica y en tiempo real para la prevención y atención médica.

_Generación de scores personalizados: se desarrollará un algoritmo capaz de asignar scores de riesgo cardiovascular personalizados, facilitando la toma de decisiones clínicas y la priorización de intervenciones.

_Identificación de patrones específicos de riesgo en la población afiliada.

_Sistema de alerta temprana: la implementación de un sistema de alerta temprana ayudará a la detección anticipada de pacientes en riesgo, permitiendo intervenciones preventivas oportunas y optimizando la asignación de recursos.

La integración de la inteligencia artificial en la predicción del riesgo cardiovascular representa una oportunidad innovadora para la mejora de la atención en salud. La utilización del extenso set de datos de SISAO, junto con técnicas avanzadas de machine learning, permitirá desarrollar una herramienta predictiva adaptada a las características de la población afiliada a OSEP [10]. Este enfoque no solo optimiza la detección temprana y la prevención de eventos cardiovasculares, sino que también contribuye a una asignación más eficiente de los recursos sanitarios y a la implementación de estrategias de intervención personalizadas.

4. FORMACIÓN DE RECURSOS HUMANOS

Este trabajo es financiado parcialmente por la Universidad Tecnológica Nacional y corresponde a una de las Unidades de Proyecto

integrantes del PID TETEUME0008760TC, homologado y acreditado por la Universidad Tecnológica Nacional, en el cual se busca incentivar las actividades de investigación científica en docentes y estudiantes de grado, a través de equipos interdisciplinarios. En particular se trata de un equipo multidisciplinar, con amplia participación de especialistas, como se referencia en los autores. El equipo de trabajo está conformado por profesionales de la salud, especialistas en inteligencia artificial y personal de TICs de OSEP.

5. BIBLIOGRAFÍA

[1] Framingham Heart Study. National Heart, Lung, and Blood Institute.

[2] WHO Global Status Report on Noncommunicable Diseases 2020.

[3] Documentación interna de OSEP sobre la historia clínica SISAO.

[4] D'Agostino, R. B., Vasan, R. S., Pencina, M. J., Wolf, P. A., Cobain, M., Massaro, J. M., & Kannel, W. B. (2008). *General cardiovascular risk profile for use in primary care*. Circulation, **117**(6), 743-753. https://doi.org/10.1161/CIRCULATIONAHA.107.699579

[5] Attia, Z. I., Kapa, S., Lopez-Jimenez, F., McKie, P. M., Ladewig, D. J., Satam, G., & Friedman, P. A. (2019). Screening for cardiac contractile dysfunction using an artificial intelligence–enabled ECG. *Nature Medicine*, **25**(1), 70-74. https://doi.org/10.1038/s41591-018-0240-2

[6] Rajpurkar, P., Hannun, A. Y., Haghpanahi, M., Bourn, C., & Ng, A. Y. (2017). Cardiologist-level arrhythmia detection with deep neural networks. *Nature Biomedical Engineering*, **1**, 429-436. https://doi.org/10.1038/s41551-017-0074

[7] Shickel, B., Tighe, P. J., Bihorac, A., & Rashidi, P. (2018). Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *Journal of Biomedical Informatics*, **83**, 85-100. https://doi.org/10.1016/j.jbi.2018.04.005

[8] Poplin, R., Varadarajan, A. V., Blumer, K., Liu, Y., McConnell, M. V., Corrado, G. S., Peng, L., & Webster, D. R. (2018). Predicting cardiovascular risk factors from retinal fundus photographs using deep learning. *Nature Biomedical Engineering*, **2**(3), 158-164. https://doi.org/10.1038/s41551-018-0195-0

[9] Huawei Cloud AI Certification (HCIA-AI). (s.f.). *Huawei Cloud*. Recuperado de https://www.huaweicloud.com/intl/es-us/activity/tech_community/hcia_ia.html

[10] MindSpore: Open Source AI Framework. (s.f.). *Huawei MindSpore*. Recuperado de https://en.wikipedia.org/wiki/MindSpore

[11] Huawei PanGu: Large-Scale AI Model. (s.f.). *Huawei AI Research*. Recuperado de https://en.wikipedia.org/wiki/Huawei_PanGu

[12] Huawei Cloud Documentation Center. (s.f.). *Huawei Cloud Docs*. Recuperado de https://doc.hcs.huawei.com

INTELIGENCIA COMPUTACIONAL Y ARTIFICIAL APLICADA A CIUDADES SOSTENIBLES Y RESILIENTES

Ana Carolina Olivera [a,b], Pablo Javier Vidal[a,b], Martín Cogo-Belver[b], Gastón Ezequiel Cavallo[b], Luciano José Masuelli-Redmond[b] y Francisco Devaux[b]

[a]*Instituto Interdisciplinario de Ciencias Básicas, Consejo Nacional de Investigaciones Científicas y Técicas. Universidad Nacional de Cuyo (*ICB CONICET UNCuyo*)*
[b]*Facultad de Ingeniería. Universidad Nacional de Cuyo* (FING-UNCuyo*)*
contacto: acolivera@conicet.gov.ar,carolina.olivera@ingenieria.uncuyo.edu.ar

CONTEXTO

Esta presentación corresponde a las tareas de investigación que se llevan a cabo en el ICB CONICET UNCuyo en conjunto con la Facultad de Ingeniería de la UNCuyo en el marco del proyecto 06/B052-T1 titulado *Inteligencia computacional aplicada a problemas de optimización* (2022-2025) financiado por la UNCuyo y el PICT 2020-SERIEA-00743 titulado *Diseño y desarrollo de estrategias novedosas aplicadas a problemas de tráfico vehicular y transporte público de pasajeros urbano* financiado por la Agencia Nacional de Promoción de la Investigación, el desarrollo Tecnológico y la Innovación, Argentina.

RESUMEN

Esta línea de investigación se centra en diseñar novedosas metodologías basadas en Inteligencia Artificial combinándolas con otras técnicas a fin de abordar satisfactoriamente problemas relacionados a Ciudades como son la Movilidad Urbana, Smart Farming, Internet of Things (IoT), Desastres Naturales y su impacto. El trabajo presentado describe la línea de investigación que estamos llevando adelante y los resultados que se han obtenido hasta el momento.

Palabras clave: Inteligencia Artificial, Movilidad Urbana Sostenible, Ciudades Resilientes.

1. INTRODUCCIÓN

Nuestro grupo de investigación está formado por investigadores del Instituto Interdisciplinario de Ciencias Básicas del CONICET y la Universidad Nacional de Cuyo. Asimismo, participan del grupo docentes y alumnos de la Facultad de Ingeniería de la UNCuyo.

Ciudades Sostenibles y resilientes comprende una variedad de problemas derivados de las actividades que se desarrollan en las ciudades y de cómo son capaces de sobrellevar eventos inesperados. Nuestras investigaciones abordan distintos problemas que se manifiestan en las ciudades y los resuelven utilizando diferentes métodos.

2. LINEAS DE INVESTIGACIÓN y DESARROLLO

- Diseño de novedosas hibridaciones de técnicas de Inteligencia Computacional y Artificial.
- Aplicación de herramientas computacionales a problemas del área de Movilidad Urbana Sostenible, Smart Farming, IoT y Ciudades Resilientes.
- Análisis de la situación actual de las ciudades en lo que respecta a Movilidad Urbana sostenible.

3. RESULTADOS OBTENIDOS/ESPERADOS

En lo que respecta a las líneas de investigación se realizaron diversos desarrollos en el campo de la inteligencia computacional y artificial. Asimismo, se realizaron análisis intensivos del impacto de la movilidad urbana sostenible en las ciudades. A continuación, se muestra un resumen de cada una.

3.1 Movilidad Urbana: Tráfico Urbano

En esta línea presenta una novedosa propuesta basada en el Algoritmo de Enjambres de Partículas Multiobjetivo (AEPM) con una mirada multiobjetivo del problema considerando mejorar el flujo del tráfico vehicular y el del peatón a través de la sincronización de los semáforos. La función considera como principal componente a la velocidad de los vehículos y el tiempo de espera de los peatones hasta llegar a su destino. Se busca mejorar la experiencia de los peatones y evaluar su influencia en la movilidad al reducir el uso de vehículos en las urbes. Se utilizó como escenario el microcentro de la ciudad de Mendoza que contiene semáforos peatonales y vehiculares (ver Fig. 1).

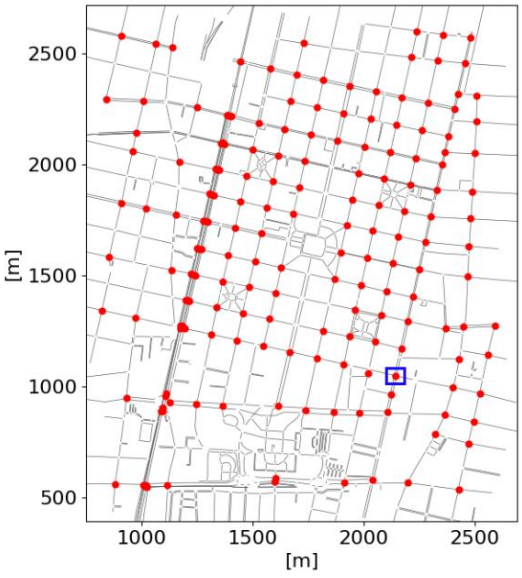


Figura 1. Ubicación de las 157 intersecciones semaforizadas sobre el centro de la ciudad de Mendoza y sus alrededores.

Los resultados demostraron que las variables incluidas en la función objetivo son mejoradas ampliamente por AEPM con respecto al resto de los algoritmos y a la configuración actual de los semáforos. Esto último, no significa que simplemente se consuma menos combustible sino también que los vehículos combustionan mejor ayudando a mantener la vida útil de los mismos y de la misma forma, los peatones pueden tener una mejora en sus tiempos promedio de traslado produciendo una mejora en la percepción de la calidad del servicio. Los resultados se encuentran publicados en la IEEE Latin American Transactions [OV22].

3.2 Movilidad Urbana Sostenible

Actualmente, existen países donde el sistema de bicicletas compartidas convive con el sistema de monopatines eléctricos compartidos que tiene similares características al de bicicletas compartidas, pero con la salvedad de que requieren cargarse para su funcionamiento. En particular, se construyó una propuesta multiobjetivo que considera la maximización del cubrimiento del sistema mientras se minimiza el costo para el operario del servicio. Se propuso una novedosa combinación de una variación del NSGA-II con programación entera mixta con el objetivo de aproximar la frontera de Pareto del problema. Los resultados de nuestra propuesta fueron comparados con otros modelos obteniendo mejoras a nivel de cubrimiento de la frontera de Pareto. Este trabajo se encuentra publicado en la prestigiosa IEEE Latin American Transactions [BO22a].

3.3 Smart Farming e Internet of Things

Se estudiaron varios algoritmos de cifrado y verificación de datos utilizando distintos conjuntos de datos y el software de criptografía OpenSSL. Las pruebas se realizan en hardware Raspberry Pi, lo que facilitó la implementación de nodos Fog. Se evalúan métricas como CPU, memoria y consumo de energía mediante scripts. Los resultados muestran que el tipo de datos no influye en la elección del algoritmo y se identifican patrones relacionados con el comportamiento del tamaño de los datos, lo que ayuda a definir estrategias de seguridad [MOVM24].

Entre otros problemas complejos abordados fue el despliegue de sensores en

agricultura. Con el objetivo de recolectar datos dentro de un área sembrada. Para esto se analizaron diferentes redes de sensores inalámbricos WSN y se comparó su desempeño con diferentes protocolos de comunicación mediante el uso de simulación [OMOV23] utilizando un caso de estudio real de la Provincia San Luis (ver Fig. 2).



Figura 2. fotos satelitales (fuente Google maps) del área simulada (Alto Pelado, San Luis, Argentina, primera imagen) y los dos escenarios definidos (segunda y tercera imagen respectivamente) [OMOV23].

3.4 Movilidad Urbana: Micromovilidad y SARS-COV2

La micromovilidad urbana ha tomado relevancia durante la pandemia en particular los sistemas monousuario como el sistema de bicicletas compartidas. Considerando la abrupta aparición del SARS-COV2, se realizó un análisis y visualización de datos de bicicletas compartidas de la Ciudad Autónoma de Buenos Aires para establecer si el virus provocado por el SARS-COV2 había tenido impacto en la movilidad urbana. Este análisis se presentó en el Décimo Segundo Encuentro de Investigadores y Docentes de Ingeniería en Los Reyunos, Mendoza [CMVO23] (ver Fig. 3).

Siguiendo con esta línea y considerando que en 2020 sobrevino de forma abrupta la pandemia derivada del SAR-COV-2 nos abocamos al problema que poseen muchos países en vías de desarrollo que no accedían en forma automática a grandes cantidades de vacunas. Se propuso un modelo de distribución de la vacuna COVID-19 y se probaron varios algoritmos de optimización. La propuesta se testeo en el escenario argentino. Presentó una importante demostración del impacto de una distribución optimizada de la vacuna en territorios como el de Argentina. Esta publicación se encuentra en publicada como capítulo de libro en la prestigiosa editorial Springer [BO22b].

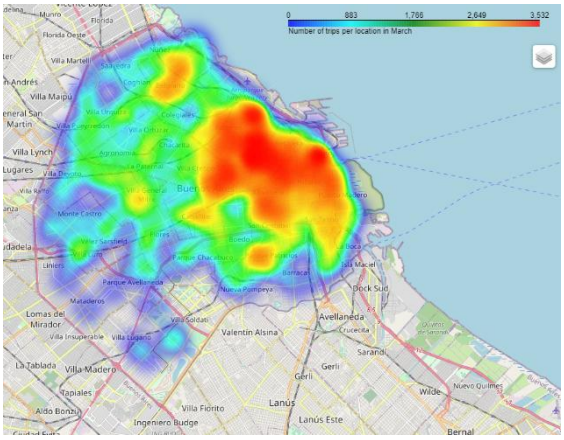


Figura 3. Volumen de viajes por cada estación de bicicletas compartidas (punto de partida) en el área de la Ciudad Autónoma de Buenos Aires en marzo del 2022 [CMVO23].

4. FORMACIÓN DE RECURSOS HUMANOS

El grupo de trabajo está formado por: Una Profesora Titular con Dedicación Semiexclusiva de la Facultad de Ingeniería de la UNCuyo (Investigadora Independiente del CONICET), Un Profesor Titular con Dedicación Simple de la Facultad de Ingeniería de la UNCuyo (Investigador Adjunto del CONICET), exBecario CIN, un exbecario de la UNCuyo, dos tesistas de grado de la Licenciatura en Ciencias de la Computación y alumnos de la Licenciatura en Ciencias de la Computación de la Facultad de Ingeniería

de la UNCuyo.

En lo que respecta a la formación de recursos humanos el Sr. Gastón Cavallo fue becario CIN durante la vigencia del proyecto y se encuentra realizando su trabajo final de carrera de la Licenciatura en Ciencias de la Computación titulado *Visualización e Interpretación de Datos de Transporte Público Urbano Mediante Inteligencia Computacional* en relación directa con los temas de los proyectos de investigación. El alumno Francisco Devaux fue becario durante la vigencia del proyecto en el tema *Detección automática de infracciones de tránsito asociadas a la falta del casco obligatorio en motos y bicicletas.* Los señores Devaux y Masuelli-Redmond se encuentran actualmente terminando un proyecto de investigación para alumnos avanzados bajo la asesoría de la Dra. Olivera que tiene como objetivo final realizar una aplicación web para cargar los videos de cámaras de vigilancia urbana para



el monitoreo de motocicletas (ver Fig. 4).

Figura 4. Frame con la detección predicha por el modelo. Es frame fue extraído de una filmación de una calle de la ciudad de Ho Chi Minh (Vietnam) a hora pico de tráfico, y está disponible en https://www.youtube.com/watch?v=1ZupwFOhjl4&ab_channel=AmazingNewLife.

El Sr. Martín Cogo-Belver está por defender tu tesina final de carrera en el tema *Evaluación de daños en las ciudades producidos por desastres naturales utilizando Machine Learning* y es dirigido por la Dra. Olivera. En su tesina el Sr. Cogo-Belver propone analizar y evaluar el desempeño de un modelo de aprendizaje profundo para la detección y clasificación de daños en estructuras urbanas a partir de imágenes satelitales (ver Fig. 5).

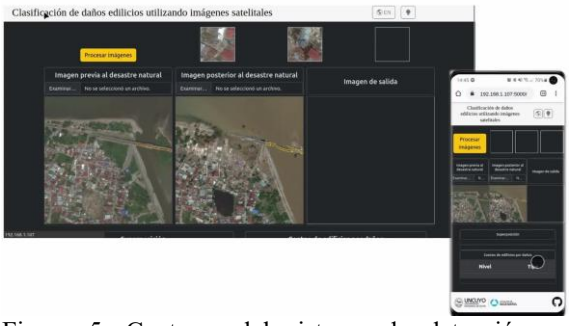


Figura 5. Capturas del sistema de detección y evaluación de daños en edificios (Cogo-Belver Tesina de grado).

5. BIBLIOGRAFIA

[BO22] Baquela, G. E. y Olivera, A. C. Humanitarian Logistics in the Disaster Risk Reduction Perspective. Springer, 2022, chapter Optimising Distribution of Limited COVID-19 Vaccines: Analysing Impact in Argentine, pp. 273–291. ISBN 978-3-030-90877-5. DOI: 10.1007/978-3-030-90877-5_9.

[BO22b] Enrique Gabriel Baquela and Ana Carolina Olivera. Optimising Distribution of Limited COVID-19 Vaccines: Analysing Impact in Argentine, pages 273–291. Springer International Publishing, Cham, 2022. URL: 10.1007/ 978- 3- 030- 90877-5_9.

[CMVO23] Gastón Ezequiel Cavallo, Emmanuel N. Millán, Pablo Javier Vidal, and Ana Carolina Olivera. Análisis del impacto del covid-19 sobre el sistema de bicicletas compartidas en la ciudad autónoma de buenos aires. In Décimo Segundo Encuentro de Investigadores y Docentes de Ingeniería, Mendoza, 2023.

[MOVM24] Layla M. Martínez Guevara, Pablo Javier Vidal, Ana Carolina Olivera, and Emmanuel N. Milán. Analysis and comparison of encryption and verification algorithms applied to a fog computing architecture. Journal of Information and Telecommunication, 0(0):1–17, 2024.URL https://doi.org/10.1080/24751839.2024.2411884

[OMOV23] Darío Maximiliano Orozco, Emmanuel N. Millán, Ana Carolina Olivera, and Pablo Javier Vidal. Modelado,

análisis y selección de tecnologías IoT para smart farming regional a través de entornos de simulación. In Décimo Segundo Encuentro de Investigadores y Docentes de Ingeniería, Mendoza, 2023.
[OV22] Ana Carolina Olivera and Pablo Javier Vidal. A multi-objective swarm approach with pedestrians spotlight in traffic urban optimisation. IEEE Latin America Transactions, 20(11):2363–2370, 2022. Scimagojr Q2. URL: http://doi.org/10.1109/TLA.2022.9904761.
[OVB+22] Ana Carolina Olivera, Pablo Javier Vidal, Enrique Gabriel Baquela, Marisa Haderne, Mariel Denise Volman Stern, and Gastón Cavallo. Inteligencia computacional aplicada a movilidad urbana. In Universidad Champagnat, editor, Libro de Actas WICC 2022, XXIV Edición del Workshop de investigadores en Ciencias de la Computación, pages 51–53, April 2022.

Machine Learning para el análisis metacognitivo de la sincronización sensorio-motora en runners

Pacheco, Marianela[1,2,3]; Guevara Cruz, Ronny S.[1,3]; Trujillo-Jiménez, M. Alexandra[1,3]; Claudio A. Delrieux[1,3]; Gustavo Gasaneo[2, 3]

[1]Laboratorio de Ciencias de las Imágenes, Instituto de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur, CONICET, 8000 Bahía Blanca, Argentina.
[2]Neufisur, Instituto de Física del Sur, Departamento de Física, Universidad Nacional del Sur, CONICET, 8000 Bahía Blanca, Argentina
[3]Universidad Provincial del Sudoeste, 8000 Bahía Blanca, Argentina.
marianelapachecomt@gmail.com

Área Temática: Agentes y Sistemas Inteligentes

Resumen

El análisis metacognitivo de la sincronización sensorio-motora permite comprender cómo los individuos que practican *running* toman registro de la regulación y ajustes de sus movimientos en respuesta a estímulos rítmicos internos y externos. Esta línea de investigación propone el uso de técnicas de *Machine Learning* (ML) para analizar la relación entre la sincronización rítmica y el rendimiento en el *running*. El objetivo es investigar y desarrollar técnicas informáticas de bajo costo que faciliten soporte en la toma de decisiones basadas en datos a los expertos en el diseño de entrenamientos deportivos. Para hacer frente a esta problemática, se plantea el desarrollo de algoritmos de aprendizaje no supervisado, específicamente clustering, y modelos predictivos para identificar patrones que permitan optimizar el rendimiento deportivo. En este estudio se analizaron autoreportes en referencia a la percepción de la sincronización sensorio-motora con la música (estímulo auditivo externo), la respiración (estímulo propioceptivo interno) y la interacción con otros corredores (*runners)* (estímulo visual externo). Los resultados preliminares sugieren que la correcta percepción de los estímulos que da lugar a la sincronización con la respiración influye significativamente en el rendimiento de quien corre. El impacto de la música y de la presencia de otros corredores, por otro lado, varía de acuerdo al perfil de la persona.

Contexto

El presente trabajo se lleva a cabo en el Instituto de Ciencias e Ingeniería de la Computación (ICIC, UNS), en el marco del Consejo Nacional de Investigaciones Técnicas y Científicas (CONICET), y en colaboración con el Instituto Físico del Sur (IFISUR) y el Laboratorio de Ciencias de las Imágenes (LCI). Los trabajos realizados en esta línea implican la participación multidisciplinaria de docentes, investigadores, y becarios doctorales y posdoctorales.

1. Introducción

El análisis metacognitivo de la sincronización sensorio-motora es un campo emergente que explora cómo los individuos regulan y ajustan conscientemente el ritmo de sus movimientos en función de estímulos externos e internos, tales como la música, la respiración y la interacción con otros corredores. Comprender estos procesos no solo es clave para optimizar

el rendimiento deportivo, sino que también aporta información valiosa sobre el control motor y la adaptación cognitiva en actividades físicas.

Tradicionalmente, el estudio del rendimiento deportivo ha sido abordado desde una perspectiva biomecánica y fisiológica (Brick, N. E., Campbell, M. J., Sheehan, R. B., Fitzpatrick, B. L., & MacIntyre, T. E., 2018). La incorporación de modelos computacionales y herramientas de ML permite analizar perfiles de comportamientos que podrían facilitar la toma de decisiones basada en datos (Wang et al., 2021). La sincronización sensorio-motora es un fenómeno clave en deportes como el running, donde la coordinación del ritmo de los movimientos realizados con estímulos externos e internos puede optimizar el rendimiento y reducir la fatiga (Karageorghis et al., 2019).

En este contexto, la aplicación de herramientas de informática y sistemas inteligentes permite evaluar de manera objetiva y escalable los patrones de sincronización y su impacto en el desempeño de los runners. Las técnicas de ML e Inteligencia Artificial (IA) ofrecen nuevas oportunidades para modelar y predecir estos procesos, identificando perfiles de runners que pueden beneficiarse más del uso de estímulos rítmicos en sus entrenamientos. En este trabajo, se explora la aplicación de algoritmos de clustering, análisis de componentes principales (PCA) y modelos de clasificación supervisados para comprender mejor este fenómeno y su impacto en el desempeño deportivo, más específicamente en el *running*.

2. Líneas de Investigación y Desarrollo

Los principales objetivos de este estudio en desarrollo son:

- Explorar la relación entre la percepción de la sincronización sensorio-motora y el rendimiento en el *running* mediante técnicas avanzadas de ML e IA.
- Desarrollar un modelo predictivo que genere descripciones de los perfiles de corredores y los clasifique según las características de sus prácticas deportivas. El objetivo es que el modelo permita identificar qué perfiles podrían beneficiarse de entrenamientos enfocados en el desarrollo de habilidades sensorio-motoras, evaluadas rítmicamente durante sus sesiones de entrenamiento.
- Contribuir al desarrollo de estrategias de entrenamiento personalizadas basadas en evidencia, integrando hallazgos sobre la sincronización sensorio-motora, el ritmo y el rendimiento deportivo.

Para alcanzar estos objetivos, se han implementado diversas metodologías:

- **Preprocesamiento de datos:** limpieza, transformación y normalización de las respuestas obtenidas en la encuesta aplicada a los corredores.
- **Análisis exploratorio de datos:** visualización y estadísticas descriptivas para comprender la distribución de las respuestas.
- **Clustering:** aplicación de algoritmos K-means y DBSCAN para agrupar runners según sus patrones de entrenamiento y experiencias de sincronización sensorio-motora.
- **Análisis de Componentes Principales (PCA):** reducción de dimensionalidad para identificar las variables con mayor influencia en la separación de clusters.

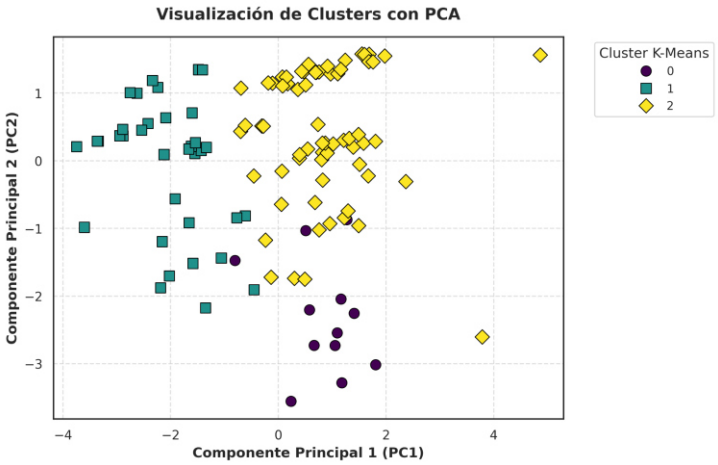
- **Visualización de datos:** representación gráfica de patrones identificados en el análisis.
- **Modelos predictivos:** implementación de clasificadores supervisados y regresión para predecir el impacto de la sincronización sensorio-motora en el rendimiento.
- **Evaluación de modelos:** validación cruzada y métricas para medir la precisión de los modelos

3. Resultados Obtenidos / Esperados

La arquitectura del modelo propone dos etapas: preprocesamiento y experimentación. La etapa de preprocesamiento se encarga de preparar la base de datos y enviarlas a los modelos de aprendizaje no supervisados. Esta, verifica los valores nulos, valores atípicos, codifica los datos mediante el método estadístico DUMMY y reduce su dimensionalidad para visualización mediante la técnica no lineal t-Distributed Stochastic Neighbor Embedding (T-SNE) (Cai & Ma, 2022). En la etapa de experimentación se aplica clustering y modelos de aprendizaje no supervisado para identificar patrones en la experiencia de sincronización y su impacto en el desempeño.

Los resultados preliminares indican:

- Identificación de tres perfiles principales de corredores según su experiencia de sincronización (ver Figura 1) estuvo determinada por un análisis de componentes principales (PCA), donde las variables más representativas en la separación fueron la participación en carreras (PC1: 0.471784), el tipo de entrenamiento realizado (PC1: 0.422958), la distancia máxima recorrida en running (PC1: 0.412431), la frecuencia semanal de entrenamiento (PC1: 0.393118) y el uso de aplicaciones para evaluar el desempeño (PC1: 0.364684). Estos factores reflejan diferencias clave en la experiencia metacognitiva de la sincronización sensorio-motora y su posible impacto en el rendimiento deportivo.
- La mayoría de los corredores reportaron haber experimentado sincronización con la respiración. No ocurrió lo mismo con las otras variables valoradas: la percepción de sincronización con la música y otros corredores. Esto sugiere que la sincronización sensorio-motora con la respiración es un fenómeno común entre corredores con distintos niveles de experiencia y motivaciones. Sin embargo, se aprecia una ligera variabilidad en la proporción de respuestas afirmativas, destacando que quienes corren por desafío personal o por razones competitivas
- **Figura 1 -** *Análisis de componentes principales.*



podrían presentar una menor tendencia a percibir esta sincronización en comparación con quienes lo hacen por salud, relajación o socialización.

El presente trabajo es un estudio preliminar que combina aspectos neuropsicológicos

involucrados en la práctica deportiva, con herramientas de ML, orientado a transformar el diseño de los entrenamientos deportivos basados en datos. Acorde a lo recién mencionado, se definen los siguientes objetivos esperados:

- Impulsar la creación de conocimiento en áreas clave para el desarrollo de tecnologías innovadoras, accesibles y de bajo costo computacional, que promuevan soluciones avanzadas.
- Fortalecer la actividad de investigación y vinculación que contribuya a mejorar prácticas basadas en evidencia en el ámbito de la psicología, específicamente, de la psicología deportiva.
- Siguiendo con la línea basada en algoritmos de Aprendizaje Supervisado y No Supervisado se espera poder predecir perfiles de corredores para identificar sus necesidades y aportar estimulación acorde a la necesidad del individuo.
- Desarrollar aplicaciones móviles que se adapten al perfil del corredor y sincronicen automáticamente la música con su cadencia de carrera, utilizando datos del acelerómetro y giroscopio del smart phone o Smart Watch, junto con técnicas de Time Stretching para ajustar el BPM de las canciones sin distorsionar su calidad.
- Desarrollar algoritmos escalables capaces de manejar grandes volúmenes de datos de manera eficiente, garantizando su desempeño y adaptabilidad del entorno y localidad de los runners.

4. Formación de Recursos Humanos

A continuación, se describen los perfiles de los participantes del estudio y sus especialidades relacionadas:

- Marianela Pacheco. Becaria doctoral del CONICET. Docente universitaria. Musicoterapeuta clínica. Tecnología de Detección de Desórdenes Sensoriales, procesamiento de datos y señales.
- Ronny S. Guevara Cruz. Becario doctoral del CONICET. Docente universitario. Ciencia de datos, procesamiento de imágenes y señales, y Machine Learning.
- M. Alexandra Trujillo Jiménez. Becaria postdoctoral del CONICET. Docente universitaria. Ciencia de datos, procesamiento avanzado de imágenes y datos, Machine Learning / Deep Learning.
- Claudio Delrieux. Investigador principal del CONICET. Docente universitario y Director del Laboratorio de Ciencias de las Imágenes del Instituto de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur.
- Gustavo Gasaneo Investigador principal del CONICET. Docente universitario y director de Neufisur, grupo de Neurociencias Aplicadas del Departamento de Física de la Universidad Nacional del Sur.

Este proyecto contribuye a la formación de becarios, investigadores y profesionales en el área de ciencia de datos aplicada, proporcionando herramientas para el análisis de datos y el desarrollo de modelos predictivos en entornos reales. Asimismo, busca fomentar la colaboración interdisciplinaria entre especialistas en informática, neurociencia y ciencias del deporte.

5. Bibliografía

Álvarez, P., Guiu, A., Beltrán, J. R., de Quirós, J. G., & Baldassarri, S. (2019, September). DJ-Running: An Emotion-based System for Recommending Spotify

Songs to Runners. In icSPORTS (pp. 55-63).

Álvarez, P., Zarazaga-Soria, F. J., & Baldassarri, S. (2020). Mobile music recommendations for runners based on location and emotions: The DJ-Running system. Pervasive and Mobile Computing, 67, 101242.

Brick, N. E., Campbell, M. J., Sheehan, R. B., Fitzpatrick, B. L., & MacIntyre, T. E. (2018). Metacognitive processes and attentional focus in recreational endurance runners. International Journal of Sport and Exercise Psychology, 1–18. https://doi.org/10.1080/1612197X.2018.1519841

Cai, T. T., & Ma, R. (2022). Theoretical Foundations of t-SNE for Visualizing High-Dimensional Clustered Data. Journal of Machine Learning Research, 23, 173-175.

Chen, Y., Chen, C. C., Tang, L. C., & Chieng, W. H. (2024). Enhancing Running Exercise With IoT, Blockchain, and Heart Rate Adaptive Running Music. IEEE Access.

Derie, R. (2022). Running on good vibes: Music induced running-style adaptations for lower impact running (Doctoral dissertation, Ghent University).

Gaussier, E., Glesser, D., Reis, V., & Trystram, D. (2015, November). Improving backfilling by using machine learning to predict running times. In Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (pp. 1-10).

Karageorghis, C. I., Hutchinson, J. C., Bigliassi, M., Watson, M. P., Perry, F. A., Burges, L. D., Melville-Griffiths, T., & Gomes-Baho, T. J. (2019). Efectos de la sincronización auditiva-motora en el rendimiento en el sprint de 400 m: un estudio aplicado. Revista Internacional de Ciencias y Entrenamiento Deportivo, 14(6), 738-748. https://doi.org/10.1177/1747954119879359

Smyth, B., Lawlor, A., Berndsen, J., & Feely, C. (2022). Recommendations for marathon runners: on the application of recommender systems and machine learning to support recreational marathon runners. User Modeling and User-Adapted Interaction, 32(5), 787-838.

Wang, S., Scheider, S., Sporrel, K., Deutekom, M., Timmer, J., & Kröse, B. (2021). What are good situations for running? A machine learning study using mobile and geographical data. Frontiers in Public Health, 8, 536370.

Métodos para el Análisis de Datos Imprecisos: Lógica Difusa y R-Shiny

Matilde Inés Césari[1], Santiago Pérez[1]

[1]GIDECo Grupo Regional UTN de Investigación y Desarrollo de Ecosistemas de Conocimiento - ReAVi Laboratorio de investigación Realidad Aumentada y Virtual) – Universidad Tecnológica Nacional FRM / matilde.cesari@frm.utn.edu.ar

RESUMEN

El análisis de datos imprecisos es un desafío común en diversas áreas de investigación y aplicación práctica. Las observaciones imprecisas pueden surgir debido a limitaciones en las herramientas de medición y la naturaleza del fenómeno estudiado. La lógica difusa proporciona un marco flexible y realista para tratar con esta imprecisión. Este artículo explora el uso de la lógica difusa en el análisis multivariado de datos imprecisos, utilizando aplicaciones Shiny de R para mejorar la calidad y fiabilidad de los análisis. La metodología propuesta incluye la codificación y estandarización de datos, la definición de variables lingüísticas difusas y el uso de métodos multivariados para el análisis de datos borrosos.

Palabras clave: lógica difusa, datos imprecisos, análisis multivariado, lenguaje R, ciencia de datos.

INTRODUCCIÓN

En un mundo impulsado por la información, comprender y manejar datos imprecisos es esencial para la toma de decisiones en diversos campos. La imprecisión es común en muchas aplicaciones, y su tratamiento adecuado es clave para extraer conocimientos valiosos. La lógica difusa es una herramienta eficaz para gestionar la incertidumbre en los datos, permitiendo decisiones más informadas. Kahraman *et al*. (2022) exploran las extensiones de la lógica difusa AHP/ANP, destacando su uso en decisiones con comparaciones imprecisas, como evaluaciones cualitativas. Qayoom y Baig (2022) proponen una medida de entropía difusa para tratar datos imprecisos, brindando una base matemática sólida para la teoría de la información difusa. Swathi M (2023) presenta una visión general de la lógica difusa y sus aplicaciones, subrayando su relevancia en inteligencia artificial y sistemas de control. Lian (2020) introduce un nuevo enfoque para el procesamiento de información imprecisa, destacando sus ventajas sobre la tecnología difusa tradicional.

En este contexto, la lógica difusa se consolida como una herramienta efectiva para representar y gestionar la incertidumbre.

La imprecisión en los datos puede deberse a varias razones, como valores perdidos, errores de medición, subjetividad y la complejidad del fenómeno estudiado. Estos factores pueden introducir sesgos y distorsiones en el análisis de datos, afectando la validez de los resultados y comprometiendo la calidad de las interpretaciones.

La lógica difusa se distingue de los métodos estadísticos tradicionales en que permite trabajar con datos que no se ajustan estrictamente a las categorías de verdadero o falso. A diferencia de los enfoques estadísticos convencionales, que requieren distribuciones de datos bien definidas y categorizaciones claras, la lógica difusa ofrece un enfoque más flexible y realista para el análisis de datos imprecisos. Esto se logra al permitir grados de pertenencia en lugar de categorizaciones binarias, facilitando así la modelización de problemas complejos donde la ambigüedad y la incertidumbre son inherentes.

La aplicación de la lógica difusa en el análisis de datos se explora en varios dominios, demostrando su flexibilidad y eficacia en el manejo de datos imprecisos e inciertos. Kwiatkowski *et al*. (2022) presentan un análisis

multidimensional basado en lógica difusa de datos de incidentes de tráfico, destacando cómo la lógica difusa puede gestionar la incertidumbre y vaguedad inherentes en los datos de tráfico reales, mejorando así el análisis de los factores que contribuyen a los accidentes de tráfico. Bolodurina y Speshilov (2023) se centran en la gestión del transporte de carga bajo incertidumbre, proponiendo un algoritmo basado en reglas de lógica difusa para optimizar la selección de rutas y los procesos de toma de decisiones, adaptándose a las preferencias del cliente y reduciendo los riesgos. Agayan *et al*. (2023) aplican la lógica difusa al análisis de series temporales, utilizándola para identificar anomalías y características morfológicas dentro de los datos. Ulumuddin (2023) introduce un algoritmo de inferencia difusa que integra bases de datos relacionales, mejorando la eficiencia de los sistemas de inferencia difusa al minimizar el uso de memoria y el tiempo de computación. Finalmente, Djurayev y Matkurbonov (2023) proponen un modelo para mejorar la eficiencia del enrutamiento en redes de transmisión de datos utilizando lógica difusa, demostrando cómo las métricas difusas pueden mejorar las decisiones de enrutamiento adaptativo.

La accesibilidad a metodologías avanzadas, como la lógica difusa, está restringida para muchos usuarios finales debido a la necesidad de un conocimiento técnico profundo. La lógica difusa es una forma de lógica multivaluada que se utiliza para manejar la incertidumbre y la imprecisión, permitiendo una representación más flexible y matizada de la realidad. Sin embargo, implementar estas técnicas requiere habilidades técnicas avanzadas, lo que limita su adopción por parte de usuarios sin formación especializada.

La falta de herramientas intuitivas y amigables para el usuario es una barrera significativa para la implementación práctica de la lógica difusa.

Esto crea un desafío para los usuarios que podrían beneficiarse de estas técnicas, pero no tienen la formación necesaria para utilizarlas eficazmente. Para mejorar la accesibilidad, se necesitan desarrollos en varias áreas. En primer lugar, es crucial crear herramientas y plataformas que simplifiquen la configuración y el uso de la lógica difusa, reduciendo la necesidad de programación compleja. Esto podría incluir interfaces gráficas de usuario (GUIs) que permitan a los usuarios construir sistemas de lógica difusa a través de interacciones visuales en lugar de código.

Este artículo propone una metodología innovadora que integra algoritmos de lógica difusa con aplicaciones Shiny de R, facilitando el análisis de datos imprecisos y promoviendo la toma de decisiones informadas. El objetivo general es ampliar el alcance de la aplicación práctica y precisión de las soluciones de la lógica difusa en el tratamiento y análisis de datos imprecisos. Para lograrlo, se implementará una metodología innovadora que integra algoritmos de lógica difusa y análisis multivariado utilizando el lenguaje R y el paquete Shiny.

METODOLOGÍA

1. Codificación y Estandarización de Datos

La primera etapa en la metodología propuesta es la codificación y estandarización de los datos imprecisos. Los datos observados se tabularon en tablas cuantitativas y cualitativas, clasificándolos en categorías numéricas y nominales, respectivamente. Para eliminar problemas de cálculo originados por la utilización de diferentes escalas, se normalizan las puntuaciones de los atributos a un rango de 0 a 1 mediante el método de ecuación de la recta.

Código 1 *Normalización de los datos*

```
# Instalación y carga de la librería 'caret'
install.packages("caret")
library(caret)

# Normalización de los datos, centrado y
escalado
datosTransf <- preProcess(datos, method =
c("center", "scale"))
)
```

2. Transformación a Datos Borrosos

El proceso de borrosificación transforma las percepciones en números difusos utilizando funciones de pertenencia. Las variables lingüísticas difusas se definen mediante descriptores lingüísticos y funciones de

pertenencia como las funciones triangular, trapezoidal y gaussiana. La elección del tipo de función de pertenencia depende de la naturaleza de los datos y el objetivo del análisis.

Definición de Variables Lingüísticas Difusas

Para representar adecuadamente las variables difusas, se seleccionan descriptores lingüísticos apropiados y se definen sus semánticas. Las funciones de pertenencia comúnmente utilizadas incluyen la función triangular, trapezoidal y gaussiana, cada una con sus propias características y aplicaciones ideales. La definición de parámetros de las funciones de membresía puede realizarse mediante métodos como la partición óptima univariada o a partir de rangos predefinidos.

Partición Óptima Univariada: Este método, desarrollado por Fisher en 1958, se utiliza para determinar los cortes que minimizan la varianza dentro de los grupos y maximizan la varianza entre grupos. Esta técnica es útil cuando no se tiene una teoría subyacente clara que guíe la segmentación de los datos.

Código 2 *Partición Óptima Univariada*

```
# Instalación del paquete classInt
install.packages("classInt")

# Carga de la librería classInt
library(classInt)

# Partición óptima univariada con el método
de Fisher
intervalos <- classIntervals(datos$variable,
n = 5, style = "fisher")
```

Nota: El código y más detalles están disponibles en el siguiente repositorio para posibles colaboraciones y futuras aplicaciones (Césari, 2024)

Rangos Predefinidos: Los rangos predefinidos se utilizan cuando se tiene conocimiento previo de los límites de los datos. Estos rangos se pueden definir de manera que representen adecuadamente la variabilidad observada en los datos.

Código 3 *Definición de Rangos*

```
# Definición de los límites
limites <- c(0, 1, 3, 5)
a <- limites[1]
b <- limites[3]
c <- limites[4]

# Función de membresía triangular
triangular <- function(x, a, b, c) {
  return(pmax(pmin((x - a) / (b - a), (c - x)
/ (c - b)), 0))
}
```

Nota: El código y más detalles están disponibles en el siguiente repositorio para posibles colaboraciones y futuras aplicaciones (Césari, 2024)

Borrosificación en la Práctica Analítica

La borrosificación convierte las valoraciones estandarizadas de los panelistas en números difusos, utilizando funciones de pertenencia definidas para transformar los valores numéricos en grados de pertenencia a conjuntos difusos. Los datos borrosos se representan en una tabla de contingencia, permitiendo un análisis más robusto y realista.

Código 4 *Transformación a Datos Borrosos*

```
# Carga de la librería FuzzyR para análisis
difuso
library(FuzzyR)

# Evaluación de las funciones de membresía
infer <- lapply(1:nrow(datosTransf),
function(i) {
  evalmf(datosTransf[i, ], calma)
})
```

Nota: El código y más detalles están disponibles en el siguiente repositorio para posibles colaboraciones y futuras aplicaciones (Césari, 2024)

3. Análisis Multivariado

Para el análisis multivariado de los datos borrosos, se utilizan métodos factoriales de correspondencias y análisis factorial múltiple. Estas técnicas permiten identificar patrones y relaciones significativas entre las variables difusas, facilitando la interpretación de los datos.

Código 5 Análisis Factorial

```
# Instalación del paquete FactoMineR
install.packages("FactoMineR")

# Carga de la librería FactoMineR
library(FactoMineR)

# Análisis Factorial Múltiple (AFM) sobre
datos transformados
res.famd <- FAMD(datosTransf, graph = FALSE)
```

Nota: El código y más detalles están disponibles en el siguiente repositorio para posibles

colaboraciones y futuras aplicaciones (César, 2024)

Tablas de Contingencia y Pruebas Estadísticas

Las tablas de contingencia se utilizan para analizar y comprender la relación entre variables categóricas. Para manejar frecuencias pequeñas o valores decimales, se pueden escalar los datos. Las pruebas estadísticas como la prueba exacta de Fisher y el valor de test se aplican para evaluar la independencia de las variables.

Código 6 *Prueba Exacta de Fisher*

```
# Creación de una tabla de contingencia para
dos variables tabla <- table(datos$variable1,
datos$variable2) # Prueba exacta de Fisher
para tablas pequeñas fisher.test(tabla)
```

Nota: El código y más detalles están disponibles en el siguiente repositorio para posibles colaboraciones y futuras aplicaciones (César, 2024)

4. Aplicaciones Shiny de R

Las aplicaciones Shiny de R permiten la interacción dinámica con los usuarios, facilitando la selección de variables, el número de intervalos y la precisión de los datos. Shiny proporciona una plataforma versátil para visualizar y analizar los datos de manera efectiva.

DISCUSIÓN

La lógica difusa se ha utilizado ampliamente en diversas áreas para manejar la incertidumbre y la vaguedad en los datos. Comparada con los métodos tradicionales, la lógica difusa ofrece una mayor flexibilidad y realismo en el análisis de datos imprecisos. En estudios previos, como los realizados por Bonissone y Decker (2013), y Espinilla *et al*. (2008), se ha demostrado que la lógica difusa mejora la interpretación de datos complejos y subjetivos.

Por ejemplo, Bonissone y Decker (2013) utilizaron lógica difusa para aplicaciones empresariales, mostrando cómo se pueden modelar situaciones con datos inciertos y vagos. Espinilla *et al*. (2008) compararon diversos métodos de fusión de datos heterogéneos en sistemas difusos, concluyendo que la lógica difusa proporciona resultados más precisos y significativos que los métodos convencionales. Bouhentala *et al.* (2019) proponen mejoras en los algoritmos de sistemas difusos que permiten una optimización más eficiente de los procesos de decisión, lo que es particularmente útil en situaciones donde la imprecisión y la incertidumbre son predominantes..

En el contexto del análisis multivariado, la lógica difusa permite manejar la complejidad de los fenómenos estudiados, como la percepción de calidad de vida o la satisfacción del cliente. La metodología propuesta en este artículo extiende estos enfoques al combinar la lógica difusa con aplicaciones Shiny de R, facilitando la visualización y el análisis interactivo de los datos imprecisos.

Las aplicaciones Shiny proporcionan una plataforma poderosa para implementar y visualizar el análisis difuso, permitiendo a los usuarios interactuar con los datos y ajustar los parámetros del análisis en tiempo real. Esto mejora significativamente la accesibilidad y la aplicabilidad de la lógica difusa en diversos contextos de investigación y práctica aplicada.

CONCLUSIONES

El uso de la lógica difusa y las aplicaciones Shiny de R en el análisis multivariado de datos imprecisos proporciona una metodología innovadora y efectiva para manejar la incertidumbre y mejorar la calidad de los análisis. Este enfoque permite representar la imprecisión de manera más realista y obtener resultados más fiables, lo que es crucial para la investigación y la toma de decisiones informadas.

REFERENCIAS

Agayan S., Kamaev D., Bogoutdinov S., Aleksanyan A., Dzeranov B. (2023). Time Series Analysis by Fuzzy Logic Methods. Algorithms, 16(5): (p. 238-238). https://doi.org/10.3390/a16050238

Bazila, Qayoom., M., A., K., Baig. (2021). Mathematical Interpretation of Fuzzy Information Model. (p. 459-466). https://doi.org/10.1007/978-981-16-1740-9_37

Bonissone, P. P., & Decker, K. S. (2013). Fuzzy logic applications in business. Communications of the ACM, 36(3), (p. 32-43). https://www.scirp.org/journal/paperinformation?paperid=104159

Bouhentala, M., Ghanai, M., & Chafaa, K. (2019). An enhanced EDA algorithm for fuzzy systems. Expert Systems with Applications, (p. 131, 157-167). https://www.sciencedirect.com/science/article/abs/pii/S0165011418303348

Cengiz, Kahraman., Selcuk, Cebi., Sezi, Cevik, Onar., Başar, Öztayşi. (2022). Recent Developments on Fuzzy AHP&ANP Under Vague and Imprecise Data. International Journal of the Analytic Hierarchy Process, 14(2) https://doi.org/10.13033/ijahp.v14i2.1033

Césari, M. (2024). DaFu: Análisis de datos imprecisos utilizando lógica difusa y Shiny. GitHub. https://github.com/matucesari/DaFu

Djurayev R., Matkurbonov D, Khojiakbar U. (2023). Analysis of a Model for Improving the Efficiency of Routing Control in Data Transmission Networks Based on Fuzzy Logic. Communications. https://doi.org/10.11648/j.com.20231101.11

Espinilla, M., Martínez, L., Pérez, A., & Liu, J. (2008). A comparative study of heterogeneous data fusion methods in fuzzy systems. Expert Systems with Applications, 34(3), (p. 2102-2112). https://www.worldscientific.com/doi/abs/10.1142/S0218488512400120

Hasbi, Ulumuddin. (2022). Fuzzy Inference Algorithm Using Databases. (p. 444-451). https://doi.org/10.1007/978-3-031-35314-7_39

Irina, Bolodurina., E., A., Speshilov. (2023). Application of fuzzy logic rules for data analysis and decision-making in cargo transportation management under conditions of uncertainty. Bulletin of the South Ural State University. Ser. Computer Technologies, Automatic Control & Radioelectronics, 23(2): (p.52-64). https://doi.org/10.14529/ctcr230205

Kwiatkowski M., Zacharias B., Leung C., Tsu P., Thomas J., Kolisnyk M., Cuzzocrea A. (2022). A Fuzzy-Logic Based Multi-Dimensional Analysis of Traffic Incident Data. (p.1-8). https://doi.org/10.1109/FUZZ-IEEE55066.2022.9882787

Shiyou, Lian. (2020). A New Theoretical and Technological System of Imprecise-Information Processing. viXra, https://arxiv.org/pdf/1610.02751

Solo, Ashu & Gupta, Madan. (2022). Fuzzy Logic Theory and Applications in Uncertainty Management of Linguistic Evaluations for Students. (p. 243-266) https://www.igi-global.com/chapter/fuzzy-logic-theory-and-applications-in-uncertainty-management-of-linguistic-evaluations-for-students/289194

Swathi C., Ebienazer J., Swathi M., Suruthipriya S. (2023). Fuzzy Logic. International journal of innovative research in information security, 09(03): (p.147-152). https://doi.org/10.26562/ijiris.2023.v0903.19

Zadeh, L. A. (1965). Fuzzy sets. Information and Control, 8(3), (p. 338-353). https://www.sciencedirect.com/science/article/pii/S001999586590241X

Modelo de Conocimiento como Servicio (KaaS) basado en Ontologías, Grafos de Conocimiento, IA y Gobernanza de Datos.

Ana Muñoz, Santiago Pérez, María Stefanoni, Daniela Carbonari, Matilde Césari, Bruno Roberti
GIDECo (Grupo de Ecosistemas de Conocimiento)
Departamento de ISI, Facultad Regional Mendoza, Universidad Tecnológica Nacional
(santiagocp,broberti)@frm.utn.edu.ar

Resumen

El **Modelo de Conocimiento como Servicio (KaaS)** basado en ontologías, grafos de conocimiento, IA y gobernanza de datos, proporciona una solución innovadora para la gestión y distribución de conocimiento estructurado en entornos complejos. El proyecto propone un modelo de tres capas que integra Gestión del Conocimiento, Inteligencia Artificial y Gobernanza de Datos, facilitando la toma de decisiones automatizada, y el acceso contextualizado al conocimiento. La metodología prevista incluye el diseño conceptual de ontologías, la construcción de grafos de conocimiento con tecnologías avanzadas, y la implementación de servicios inteligentes mediante agentes autónomos. La integración de políticas de gobernanza asegura la interoperabilidad y la seguridad de datos. Esta propuesta se destaca por su escalabilidad y flexibilidad, siendo aplicable en diversos dominios organizacionales. El personal principal relacionado a esta línea de investigación es docente de grado y posgrado de la Universidad Tecnológica Nacional, Facultad Regional Mendoza, y Directores de Tesis de Posgrado, y Tesistas de posgrado.

Palabras Claves

Conocimiento como Servicio, Ontologías y Grafos de Conocimiento, Inteligencia Artificial, Gobernanza de Datos, Gestión del Conocimiento

Contexto

El proyecto está insertado en el ámbito del Grupo UTN GIDECo (Grupo UTN de Gestión de Conocimiento), el que hace asiento en la Facultad Regional Mendoza, de la Universidad Tecnológica Nacional. El proyecto abarca las principales líneas de investigación del mismo: Modelos de gestión del conocimiento y gobernanza de datos para los procesos de enseñanza aprendizaje.

Introducción

En la era de la información, las organizaciones se enfrentan al desafío de gestionar volúmenes masivos de datos provenientes de múltiples fuentes, con el objetivo de convertirlos en conocimiento valioso para la toma de decisiones estratégicas. La **Gestión del Conocimiento (GC)** desempeña un papel crucial al abarcar tanto el conocimiento explícito —formal y documentado— como el conocimiento tácito, que reside en la experiencia y habilidades de los individuos. Sin embargo, los enfoques tradicionales de GC, basados en sistemas aislados y jerárquicos, presentan limitaciones significativas en términos de accesibilidad, actualización y escalabilidad.

En respuesta a estos desafíos, surge el **Conocimiento como Servicio (KaaS)**, un modelo emergente que ofrece el conocimiento en sí como un servicio dinámico, accesible y contextualizado, superando las barreras de los sistemas convencionales de GC. A diferencia de los modelos tradicionales, que dependen de estructuras rígidas y flujos de información unidireccionales, KaaS integra ontologías, **grafos de conocimiento**, e **inteligencia artificial (IA),** para crear un ecosistema de conocimiento interconectado y en evolución continua. Esto permite no solo la gestión eficiente del conocimiento explícito, sino también la captura y transferencia del conocimiento tácito a través de servicios inteligentes y colaborativos.

El potencial de KaaS va más allá de la simple distribución de información; posibilita la automatización de la toma de decisiones basada en conocimiento contextualizado y la personalización de servicios a las necesidades específicas de los usuarios. Además, su integración con **gobernanza de datos** garantiza la seguridad, privacidad e interoperabilidad en entornos complejos y distribuidos, como la computación en la nube y el **Internet Industrial de las Cosas (IIoT)**.

Este proyecto propone un **Modelo de Conocimiento como Servicio (KaaS)** de tres capas que combina **Gestión del Conocimiento**, **Inteligencia Artificial** y **Gobernanza de Datos**. La metodología incluye el diseño conceptual de ontologías, la construcción de grafos de conocimiento mediante tecnologías avanzadas (como **Neo4j**, **GraphDB** y **OWL API**), y la implementación de servicios inteligentes a través de agentes autónomos. La propuesta se destaca por su escalabilidad, flexibilidad y capacidad de integración, aportando una solución innovadora a la gestión del conocimiento en organizaciones orientadas a la innovación y competitividad.

Con esta investigación, se busca no solo enriquecer el estado del arte en KaaS, sino también proporcionar una base sólida para la aplicación práctica en diversos dominios, desde la fabricación inteligente hasta la gestión del conocimiento organizacional.

Estado del Arte

La gestión del conocimiento abarca tanto el conocimiento explícito como el tácito dentro de una organización, facilitando su captura, estructuración, distribución y aplicación [1]. El conocimiento explícito es formal, documentado y fácilmente transmisible a través de bases de datos, manuales, informes y ontologías. En contraste, el conocimiento tácito reside en la experiencia, intuición y habilidades de los individuos, siendo más difícil de codificar y compartir. Además, la gestión del conocimiento integra el conocimiento embebido en procesos, reglas y rutinas organizacionales, así como el conocimiento social que emerge de la colaboración y el intercambio entre personas [2]. A través de herramientas tecnológicas como bases de datos, sistemas de gestión del conocimiento, inteligencia artificial y grafos de conocimiento, se busca transformar estos distintos tipos de conocimiento en un recurso estratégico que potencie la innovación, la toma de decisiones y la competitividad organizacional.

La metodología convencional empleada en la GC ayuda a realizar de forma organizada las operaciones de: (i) Recopilación de la información necesaria, (ii) Evaluación de la información recopilada, (iii) Puesta en común de la información dentro de la empresa o entre empresas, y (iv) Evaluación de los datos disponibles para obtener la solución adecuada para una tarea elegida [3].

El conocimiento como servicio (KaaS) es un modelo emergente donde el conocimiento se ofrece como un servicio a los consumidores. Este modelo ha ganado atención en diversos campos, desde la fabricación inteligente hasta la gestión del conocimiento organizacional [4]**.** KaaS implica aprovechar métodos avanzados de gestión del conocimiento para generar valor al proporcionar el conocimiento en sí como un servicio, en lugar de solo datos o información. KaaS integra la gestión del conocimiento (KM), una organización del conocimiento y los mercados del conocimiento [5].

Hay bastante actividad en esta línea de investigación. En [4] se explora cómo lograr el "Conocimiento como Servicio" (KaaS) en la fabricación inteligente impulsada por el Internet Industrial de las Cosas (IIoT). El trabajo destaca que, si bien el IIoT ofrece herramientas para optimizar la fabricación inteligente, la capacidad de servicio se ve limitada por la naturaleza reactiva de los servicios convencionales. En [5] se presenta un modelo de Sistema de Gestión del Conocimiento (KMS) diseñado para facilitar

el Conocimiento como Servicio (KaaS) en un entorno de computación en la nube. El sistema permite que las comunidades de práctica (CoP) accedan y utilicen el conocimiento proporcionado como salida organizacional (por ejemplo, asesoramiento) de manera eficiente, para referenciar las mejores prácticas actuales y las lecciones aprendidas, especialmente en relación con el entorno de la computación en la nube. Y en [6] se elabora un modelo de conocimiento móvil como servicio (mKaaS) en un entorno de computación en la nube (CC), en el que unen tres aspectos del conocimiento: gestión (KM), mercados y organización.

Objetivos

Objetivo General

El presente trabajo tiene como objetivo d**efinir un Modelo de Conocimiento como Servicio (KaaS)** de tres capas que integre **ontologías**, **grafos de conocimiento**, **inteligencia artificial (IA)** y **gobernanza de datos**, proporcionando una solución escalable y flexible para la gestión del conocimiento en entornos complejos.

Objetivos específicos

- Automatizar la Toma de Decisiones Basada en Conocimiento**.**
- Garantizar la Seguridad y Gobernanza del Conocimiento**.**
- Proporcionar una Base Teórica para el Desarrollo de KaaS**.**

Metodología

La metodología propuesta para el desarrollo del modelo de tres capas basado en ontologías está diseñada para permitir un desarrollo estructurado y escalable de una plataforma de conocimiento como servicio, integrando los dominios de Gestión del Conocimiento, Ontologías, Grafos de Conocimiento, Inteligencia Artificial y Gobernanza de Datos. La metodología incluye las siguientes fases:

1) La primera fase comprende el **Análisis y Diseño Conceptual**, donde se definen los conceptos clave, relaciones y servicios de la plataforma. Esta fase inicia con la identificación de requerimientos, así como la determinación de servicios autónomos. Posteriormente, se realiza el diseño de ontologías base. Este proceso se apoya en herramientas como Protégé para el modelado OWL y BPMN para el modelado de procesos.
2) La segunda fase se centra en la **Implementación de la Capa de Datos y Conocimiento**, donde se construyen la base de datos semántica y el grafo de conocimiento. Esta fase comienza con la integración de datos desde fuentes externas, luego, se procede a la creación del grafo de conocimiento, definiendo nodos y relaciones. La fase culmina con la implementación de razonadores y consultas, configurando razonadores OWL para inferencias y creando consultas SPARQL y Cypher para recuperar información, utilizando tecnologías como Neo4j, GraphDB, Apache Jena y OWL API.
3) La tercera fase aborda la **Implementación de la Capa de Aplicación y Servicios Inteligentes**, construyendo agentes autónomos que utilicen el conocimiento inferido. Esta fase incluye el desarrollo de agentes inteligentes y la implementación del motor de reglas y la construcción de API para servicios, creando interfaces REST o GraphQL e implementando un chatbot educativo basado en las ontologías.
4) La cuarta fase se enfoca en la **Integración y Gobernanza de Datos**, asegurando la interoperabilidad y la gobernanza efectiva. Esta fase incluye la definición de políticas de gobernanza, estableciendo control de acceso y permisos para distintos roles, así como la trazabilidad y auditoría de datos. También aborda la interoperabilidad con otras plataformas, integrándose con sistemas LMS mediante API y utilizando estándares como SCORM y LTI para garantizar la compatibilidad.
5) Finalmente, la quinta fase comprende la **Evaluación y Mejora Continua**, donde se optimiza el sistema mediante pruebas y retroalimentación. Esta fase incluye la

evaluación del rendimiento del grafo de conocimiento, midiendo tiempos de inferencia y recuperación de datos, y optimizando reglas y consultas SPARQL. También abarca el monitoreo del uso de servicios inteligentes, analizando las interacciones de los usuarios con los agentes y ajustando reglas y modelos según los patrones de uso identificados.

El modelo de conocimiento propuesto [7] evoluciona y se representa a través de tres capas: Capa de Aplicación e Inteligencia, Capa Semántica y de Conocimiento y Capa de Datos, como se muestra en la Figura 1. Las ontologías se utilizan como elemento de representación semántica del conocimiento. El desarrollo de un modelo por capas basado en ontologías permite una estructuración clara, escalable e interoperable de los diferentes componentes de una plataforma educativa inteligente como servicio. Este enfoque permite la separación de responsabilidades, la mejora en la toma de decisiones basada en datos y la optimización de la personalización del aprendizaje. El modelo que se está implementando proporciona una arquitectura bien organizada.

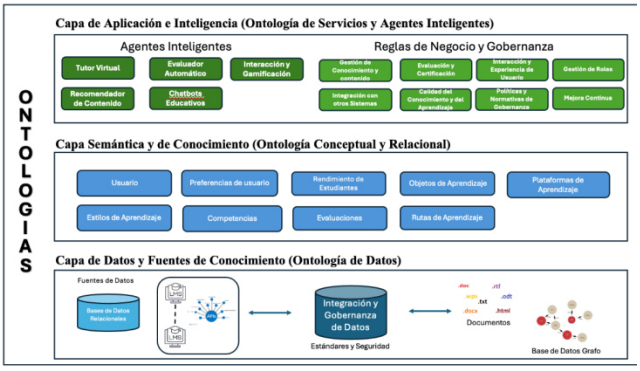


Figura 1. Modelo de capas para una Plataforma de Conocimiento como Servicio de eLearning.

Formación de RRHH

El equipo de trabajo está integrado por docentes investigadores, becarios graduados y alumnos del GIDECo de la UTN, y docentes-investigadores del CEMISID de la Universidad de Los Andes, Mérida, Venezuela. Entre estos se encuentra una tesista de Maestría, que presentará su tesis de Maestría afín a este proyecto y una tesista doctoral. Las actividades se llevan a cabo en el ámbito de las instalaciones de dichos entes, que cuentan con sus propias áreas de trabajo.

Referencias

[1] Nonaka, I. (2009). The knowledge-creating company. In *The economic impact of knowledge* (pp. 175-187). Routledge

[2] Davenport, T. H. (1998). Working knowledge: How organizations manage what they know. *NewYork Harvard Business School*.

[3] Majumder, S., & Dey, N. (2022). Artificial intelligence and knowledge management. In *AI-empowered Knowledge Management* (pp. 85-100). Singapore: Springer Singapore.

[4] Lyu, M., Li, X., & Chen, C. H. (2022). Achieving Knowledge-as-a-Service in IIoT-driven smart manufacturing: A crowdsourcing-based continuous enrichment method for Industrial Knowledge Graph. *Advanced Engineering Informatics*, *51*, 101494

[5] Abdullah, R., Eri, Z. D., & Talib, A. M. (2011, November). A model of knowledge management system for facilitating knowledge as a service (KaaS) in cloud computing environment. In *2011 International Conference on Research and Innovation in Information Systems* (pp. 1-4). IEEE

[6] Abdullah, R., & Alsharaei, Y. A. (2016, August). A Mobile Knowledge as a service (mKaaS) model of knowledge management system in facilitating knowledge sharing of cloud education community environment. In 2016 Third International Conference on Information Retrieval and Knowledge Management (CAMP) (pp. 143-148). IEEE.

[7] Muñoz, A. et al. (2023). Ambiente de Aprendizaje Inteligente desde la Gestión del Conocimiento. Simposio Internacional STEM Miami. Broward International University

POOL DE TECNOLOGÍAS 4.0 PARA IMPULSAR ORGANIZACIONES REGIONALES

Romagnano, M.[1,2,3], García, A.[2], Mejibar, D.[2], Lépez, H.[1], Aballay, A.[1], Sarmiento, A.[1], Herrera, M.[1,3], Becerra, M.[1], Ganga, L.[1,3]

[1]Departamento de Informática, FCEFN, Universidad Nacional de San Juan
[2]Instituto de Informática, FCEFN, Universidad Nacional de San Juan
[3]Departamento de Matemática, Física y Química, FCEFN, Universidad Nacional de San Juan

{maritaroma, mherrera}@iinfo.unsj.edu.ar, {dmejibar, mcbecerra2008, prof.alicia, garcia.exe, leonelganga}@gmail.com, { lepezhr, adriva2005}@yahoo.com.ar

RESUMEN

Desde hace varios años las organizaciones vienen enfrentando desafiantes maneras de manipular la información y los datos que surgen de sus transacciones diarias. Así, deben almacenar y gestionar grandes volúmenes de datos, adecuarse rápidamente al medio y tomar decisiones de forma casi inmediata.

El dinamismo del contexto en el cual están inmersas y el Big Data han dado lugar a un nuevo tipo de organizaciones; las denominadas Organizaciones Inteligentes.

Estas organizaciones se diferencian de las convencionales en cuanto a que tienen una visión integral, se apoyan en tecnologías, se basan en datos, valoran el feedback, y retienen talentos. El factor determinante en estas organizaciones es la tecnología. La capacidad de digitalizar o directamente automatizar procesos es condición primordial. Además, deben contar con una infraestructura que facilite la recolección, recopilación y análisis de datos macro.

Por ello, diversas herramientas, técnicas y metodologías, conocidas como Tecnologías 4.0, ofrecen ventajas competitivas en cuanto a la trazabilidad y el análisis de datos. Permiten conocer el estado actual y pronosticar el futuro comportamiento de las organizaciones y del entorno. Realizan acciones proactivas basadas en el análisis predictivo y prescriptivo. La rápida expansión de la Industria 4.0 o Tecnologías 4.0 gana tracción y celeridad. El mundo organizacional se redefine, se rediseña y se reinventa. Actualmente, en las organizaciones existe una delgada línea entre lo digital y lo físico. Ante esta realidad, es esencial que las organizaciones estén preparadas y alineadas con las tecnologías de punta.

Sin embargo, en la región cuyana, principalmente en San Juan, se presenta una realidad distinta. La mayoría de las organizaciones, si bien han entendido la importancia y beneficios de adoptar este nuevo paradigma tecnológico, aún se encuentran en transición y operando con tecnologías anticuadas.

Todavía queda mucho por hacer para sensibilizar sobre las ventajas competitivas de su implementación, y más en un ámbito en el que no abundan los ejemplos exitosos que impulsen al resto; siendo el primer paso para encarar los desafíos tecnológicos.

Este trabajo propone identificar un adecuado conjunto de Tecnologías 4.0 a implementar en organizaciones regionales, para asistirlas en la toma de certeras y oportunas decisiones; convirtiéndolas así en organizaciones de vanguardia, es decir en Organizaciones Inteligentes.

Palabras Claves: Analítica de Negocios, Inteligencia de Negocios, Inteligencia Artificial, Toma de Decisiones, Organizaciones Inteligentes.

CONTEXTO

El presente trabajo se encuentra enmarcado en el proyecto “Tecnologías 4.0 para Asistir a la

Toma de Decisiones en Organizaciones Regionales", presentado en la convocatoria del Consejo de Investigaciones Científicas y Técnicas y de Creación Artística (CICITCA) de la Universidad Nacional de San Juan (UNSJ), para ser desarrollado durante el período comprendido entre 01/01/2023 al 31/12/2025.

Las tareas de investigación se desarrollan en el Laboratorio de Sistemas de Información, en el ámbito del Instituto de Informática de la Facultad de Ciencias Exactas, Físicas y Naturales, UNSJ.

El grupo de investigación se encuentra conformado por profesionales y estudiantes avanzados de grado y postgrado, de distintas disciplinas, tales como: Informática, Bioingeniería, Estadística, Administración de Empresas y Abogacía. Los profesionales cuentan con una experiencia de más de 20 años en la línea de investigación, trabajando en temáticas relacionadas con inteligencia de negocios, analítica de negocios, análisis y preservación de datos, sistemas de información y recuperación de información integrando técnicas de aprendizaje automático de la Inteligencia Artificial.

1. INTRODUCCIÓN

En Latinoamérica, la mayoría las organizaciones carecen de nuevas tecnologías y la falta de digitalización de la información dificulta la obtención de datos que asistan a la toma de decisiones. Mas aún, a diferencia de los países desarrollados, hay desinformación sobre las diversas aplicaciones y lo que implica el uso de Inteligencia Artificial (AI) en el entorno organizacional (Pérez León y Rojas Arévalo, 2019).

Pangol Lascano (2022) sostiene que los países de América Latina están avanzando en este aspecto, mejorando tanto su tecnología como sus políticas públicas y estrategias de desarrollo. Sin embargo, varios países han encontrado dificultades para el desarrollo e implementación de Tecnologías 4.0, principalmente debido a la falta de inversión y de desarrollo de productos inteligentes.

En los tiempos que corren, el crecimiento de las organizaciones se sustenta en la productividad y el desempeño individual de sus colaboradores, y de todas aquellas tecnologías como la AI; útiles para trabajar de la mejor manera posible (Wang et al., 2023). La AI va más allá de ser simplemente una nueva tecnología; es un modelo único que combina capital y trabajo para impulsar una economía sostenible y sustentable. Al implementar tecnologías modernas se busca aumentar la productividad y mejorar la calidad de vida de la sociedad. Gracias a la AI, se puede crear un entorno más integrado a nivel regional y transformar la forma en que se produce el comercio internacional de bienes y servicios, adaptándose a las demandas de la globalización en constante evolución y cada vez más eficiente. Además, un aspecto muy importante, es que la AI ha cobrado mayor importancia para la toma de decisiones y la dirección estratégica (Baquero, et al., 2020).

Por ello, se viene hablando de Organizaciones Inteligentes, como un modelo relativamente novedoso. Sin embargo, ya en 1990 Peter Senge introdujo el concepto al sostener que toda agrupación tiene la posibilidad (y debería) aprender de sus propias experiencias, recopilando y analizando datos, para poder generar lo que se conoce como mejora continua. Big Data, la Inteligencia Artificial y el Machine Learning son los principales mecanismos que entran en acción en estas organizaciones (León M., Tejada G., Yataco T., 2003). Por lo tanto, estas organizaciones trabajan bajo la concepción de Paradigma 4.0 al introducir una nueva forma de producir con tecnologías de última generación, Tecnologías 4.0, para automatizar sus procesos productivos y transformar los modelos de negocio, optimizando los recursos (Ministerio de Desarrollo Productivo, 2021).

Estas tecnologías emergen a partir de la Revolución Digital, donde algunas ciencias y disciplinas como la Ciencia de los Datos (DS), AI, Inteligencia de Negocios (BI), Minería de Datos (DM), Big Data (BD), Analítica de Negocios (BA), Ingeniería del Conocimiento

(KE) y Sistemas de Información (IS) han fusionado sus esfuerzos y saberes.

Dentro del paquete de Tecnologías 4.0 se identifica la categoría de tecnologías destinadas a los sistemas de gestión organizacional (Ministerio de Desarrollo Productivo, 2021; Baruj, 2023). Así, la BI, considerada como complementaria a la DS y a la AI para el Negocio, viene a ofrecer esta gestión del conocimiento. Se presenta como un conjunto de procesos, aplicaciones y tecnologías que facilitan la obtención rápida y sencilla de datos en grandes bases de datos (BD), provenientes de distintos Sistemas de Información Organizacionales (OIS). Por lo tanto, los datos analizados e interpretados, se transforman en conocimiento apropiado para la toma de decisiones. Las posibilidades de procesamiento que ofrece BI hacen que las organizaciones puedan desempeñarse y responder dinámicamente en un entorno universal altamente impetuoso, el cual exige una constante toma de decisiones para mantenerse en el medio. Como complemento a la BI, las organizaciones deben tener en cuenta a la BA, considerada como el conjunto de técnicas y procesos tales como análisis cuantitativo, pronósticos, análisis predictivos, optimización, entre otros, que mantienen y sustentan la performance de negocios, explorando grandes volúmenes de datos para asistir a los ejecutivos en la toma de decisiones (Orrillo, 2018).

Dir&Ge (2020), pronosticaron que entre el año 2020 y el 2025, el mercado de BI crecería a ritmos anuales del 7,6%. No obstante, una de las principales restricciones para su implementación es la disponibilidad de personal capacitado. Existe amplio consenso, tanto desde el sector público como privado, respecto de la necesidad de promover la formación en habilidades en este nuevo paradigma 4.0.

La tarea de sensibilización es necesaria, y más en un contexto en el que no abundan los ejemplos exitosos que impulsen al resto; siendo el primer paso para encarar los desafíos tecnológicos.

Por lo tanto, a través de este estudio, se propone impartir conocimiento y, luego, proveer de un conjunto de técnicas, metodologías, software libre y de código abierto que ayude a las entidades regionales en su transición tecnológica y cultural, tendiendo a convertirlas en Organizaciones Inteligentes, que tomen certeras y oportunas decisiones.

2. LINEAS DE INVESTIGACIÓN Y DESARROLLO

Las tareas de investigación y desarrollo se han divido en subgrupos, trabajando en las siguientes líneas de investigación y desarrollo:

- Sistemas de información organizacionales.
- Revolución 4.0 y Tecnologías emergentes.
- Inteligencia artificial aplicada a la industria y al negocio.
- Inteligencia y analítica del negocio.
- Análisis y procesamiento estadístico de los datos.
- Preservación y legislación de los datos.

Consecuentemente, el equipo de investigación, considera la temática de relevancia social y organizacional. Posee implicancias prácticas, es viable y propicia puesto que, en abril del 2021 se lanzó el Plan de Desarrollo Productivo Argentina 4.0. Por medio de este plan se establecen estrategias para impulsar el paradigma 4.0 y promover el desarrollo de soluciones tecnológicas 4.0 (Ministerio de Desarrollo Productivo, 2021). También, se suma el plan de transformación digital (Ministerio de Obras Públicas, 2022). Además, desde el año 2020, Argentina cuenta con políticas de estado tales como Argentina Innovadora 2020 y 2030, Ley de Promoción de la Economía del Conocimiento, Industria 4.0, Plan de Inteligencia Artificial, entre otras (Ministerio de Ciencia, Tecnología e Innovación, 2020; Ministerio de Desarrollo Productivo, 2021; ARGENIA, 2020; Ministerio de Economía, 2020).

Precisamente, San Juan, la provincia donde residen los investigadores, ya viene trabajando

con polos tecnológicos como Casetic San Juan, Servicios Mineros y San Juan TEC; a la cual en el 2024 se ha sumado el Centro de Economía del Conocimiento e Innovación (CECI). Estas iniciativas se tomaron como puntapié inicial para trabajar, conjuntamente con el Estado y con organizaciones privadas, en la sustentabilidad regional, aportando conocimiento y casos de estudio desde la academia.

3. RESULTADOS OBTENIDOS

De las tareas propuestas en el plan de trabajo del proyecto, la mayoría se han completado en tiempo y forma. A continuación, se presenta un detalle de las mismas:

1. *Estudio y análisis de las distintas Tecnologías 4.0*. Se hizo una exhaustiva revisión de las tecnologías de vanguardia, principalmente aquellas relacionadas con la inteligencia y analítica de datos. Se identificó un conjunto de metodologías, procesos, aplicaciones y tecnologías que facilitan la obtención rápida y sencilla de datos en grandes DB, provenientes de distintos OIS. Así, los datos que se examinan y dilucidan, se transforman en conocimiento apropiado que sirven de suministro a la toma de decisiones.
2. *Selección de Tecnologías 4.0 que pueden incorporarse a las organizaciones regionales*. Para ello se siguieron las pautas establecidas por Gartner, quien anualmente pone a disposición un resumen gráfico: el Cuadrante Mágico de Gartner[1], de aquellas tecnologías que se ubican como líderes, retadores o aspirantes, jugadores del mercado y visionarios. Este informe anual, generado a partir de datos cualitativos, tiene como objetivo brindar una visión objetiva de la posición de cada plataforma tecnológica que ofrece inteligencia y analítica de datos.
3. *Elaboración del conjunto de técnicas, metodologías, software libre y de código abierto que se brindarán a las organizaciones regionales.* Para llevar a cabo esta tarea, se realizó un análisis comparativo de aquellas técnicas, metodologías y plataformas tecnológicas, ubicadas en el cuadrante identificado como líder, según Gartner. Luego, se realizó un cuadro comparativo de las mismas para ofrecer un análisis FODA a las organizaciones, que forman parte del estudio.
4. *Implementación de casos de estudio (pruebas piloto) en organizaciones regionales*. Para llevar a cabo esta tarea, se continúa trabajando en diversas áreas tales como: una obra social prepaga, una empresa de transporte de larga distancia y una entidad financiera. Durante todo el 2024, a través de las tareas de investigación de un alumno que está realizando su trabajo final de carrera de grado, se avanzó ampliamente con las pruebas en una organización educativa de gestión privada, y a fines de noviembre del 2024 se comenzaron tratativas con una dependencia estatal, encargada de la gestión y desarrollo de recursos hídricos de San Juan.
5. *Redefinición y ajuste del conjunto de Tecnologías 4.0*. A partir de los resultados obtenidos en las tareas 3 y 4, se comenzó a redefinir el conjunto de tecnologías a proponer a las organizaciones, principalmente considerando el contexto y los recursos disponibles.

4. FORMACIÓN DE RECURSOS HUMANOS

Dentro del grupo de investigación se identifican integrantes formándose en:

- Doctorado en Ingeniería (Facultad de Ingeniería, UNCuyo).

[1] https://www.gartner.es/es/metodologias/magic-quadrant

- Doctorado en Demografía (Escuela de graduados de la Facultad de Ciencias Económicas, UNC).
- Maestría en Estadística Aplicada (Escuela de graduados de la Facultad de Ciencias Económicas, UNC).
- Licenciatura en Ciencias de la Computación (Facultad de Ciencias Exactas, Físicas y Naturales, UNSJ).
- Licenciatura en Licenciatura en Sistemas de Información (Facultad de Ciencias Exactas, Físicas y Naturales, UNSJ).
- Beca EV CIN (Beca de Estímulo a las Vocaciones Científicas).

5. BIBLIOGRAFÍA

ARGENIA (2020). Plan Nacional de Inteligencia Artificial. Recuperado de: https://ia-latam.com/wp-content/uploads/2020/09/Plan-Nacional-de-Inteligencia-Artificial.pdf, septiembre, 2022.

Baquero, R., Moreno, M., & Tapia, I. (2020). Procedimiento para la medición y mejora de la satisfacción de los clientes. Aplicación en la empresa de Seguridad Integral sucursal Seisa Holguín. Desarrollo Sustentable, Negocios, Emprendimiento y Educación,2(5), 1-17.

Baruj, G. (2023). Tecnologías para la transformación digital en la industria argentina. 1a ed.- Ciudad Autónoma de Buenos Aires. Editorial CIECTI. Recuperado de: https://www.ciecti.org.ar/wp-content/uploads/2023/12/TTDA_book_V04.pdf, febrero del 2025.

Dir&Ge (2020). El mercado de business intelligence crecerá a ritmos anuales del 7,6% entre 2020 y 2025. Recuperado de: https://directivosygerentes.es/innovacion/mercado-business-intelligence-crecimiento, enero de 2021.

León M., Tejada G., Yataco T. (2003). Las Organizaciones Inteligentes. Notas Científicas, Vol. (6) 2: pp. 82-87. Recuperado de: https://sisbib.unmsm.edu.pe/bibvirtualdata/publicaciones/indata/Vol6_n2/pdf/organizaciones.pdf, diciembre del 2023.

Ministerio de Ciencia, Tecnología e Innovación (2020). Argentina Innovadora 2020. Principales Ejes de Políticas Públicas en CTI, https://www.argentina.gob.ar/ciencia/argentina-innovadora-2030/plan-argentina-innovadora-2020, recuperado septiembre, 2022.

Ministerio de Desarrollo Productivo (2021). Plan de Desarrollo Productivo. Argentina 4.0. https://www.argentina.gob.ar/sites/default/files/plan_de_desarrollo_productivo_argentina_4.0.vf__1.pdf, recuperado enero, 2023.

Ministerio de Economía (2020). Régimen de Promoción de la Economía del Conocimiento. Recuperado de: https://www.argentina.gob.ar/economia/igualdadygenero/regimen-de-promocion-de-la-economia-del-conocimiento#:~:text=7%2F10%2F2020, septiembre del 2022.

Ministerio de Obras Públicas (2022). Plan Estratégico de Transformación Digital para la Gestión de la Obra Pública. Recuperado de: https://www.argentina.gob.ar/obras-publicas/secretaria-gestion/transformacion-digital#:~:text=El%20Plan%20Estrat%C3%A9gico%20de%20Transformaci%C3%B3n,eficiente%20la%20Obra%20P%C3%BAblica%20Nacional, octubre del 2022.

Orrillo, J. (2018). Analítica de negocios: Una estrategia basada en la Información. Recuperado de: https://www.esan.edu.pe/conexion/actualidad/2018/03/22/analitica-negocios-informacion/, julio del 2023.

Pangol Lascano, A. (2022). Industria 4.0, implicaciones, certezas y dudas en el mundo laboral. Revista Universidad y Sociedad, 14(4), 453-465. Recuperado de: http://scielo.sld.cu/pdf/rus/v14n4/2218-3620-rus-14-04-453.pdf, diciembre del 2024.

Pérez León, E. V., & Rojas Arevalo, D. I. (2019). Impacto de la inteligencia artificial en las empresas con un enfoque global. Universidad Peruana de Ciencias Aplicadas (UPC), Lima, Perú. Recuperado de: http://hdl.handle.net/10757/628123, septiembre del 2024.

Wang, X., Lin, X. y Shao, B. (2023). Artificial intelligence changes the way we work: A close look at innovating with chatbots.*Journal ofthe Association for Information Science &Technology*,74(3), 339–353. Recuperado de: https://doi.org/10.1002/asi.24621, febrero del 2025.

PROGRAMACION BASADA EN AGENTES PARA LA SIMULACION

Monetti Julio[a], Bianchini Germán[b], Caymes-Scutari Paola[b,c]

[a]Laboratorio de Integración de Tecnologías Aplicadas a Propotipos de Software, Departamento de Ingeniería en Sistemas de Información. Universidad Tecnológica Nacional, FRM. Rodriguez 273, (M5502AJE) Mendoza.

[b]Laboratorio de Investigación en Cómputo Paralelo/Distribuido, Departamento de Ingeniería en Sistemas de Información. Universidad Tecnológica Nacional, FRM. Rodriguez 273, (M5502AJE) Mendoza

[c]Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)

jmonetti@frm.utn.edu.ar

RESUMEN

Los sistemas complejos, como el tránsito urbano y los ecosistemas, entre otros, presentan comportamientos emergentes que resultan muy dificultosos para ser explicados apropiadamente mediante enfoques tradicionales. En las últimas décadas, debido en gran parte al desarrollo del microprocesador, y arquitecturas computacionales superiores, han surgido técnicas innovadoras en el modelado de sistemas complejos. Una de estas técnicas es la simulación basada en agentes. La simulación basada en agentes ofrece una herramienta poderosa para capturar estas dinámicas al modelar cada componente del sistema como una entidad con reglas y objetivos específicos.

En el presente trabajo se aplican técnicas de modelado basado en agentes para la simulación de tránsito vehicular. Para tal fin se desarrolla una aplicación en lenguaje Java basada en autómatas celulares y agentes, con lo cual se pueden analizar comportamientos complejos de un sistema, surgidos a través de la aplicación de procesos computacionales básicos.

Palabras Clave: agentes inteligentes - autómata celular – microsimulación.

CONTEXTO

La presente línea de trabajo se encuentra inserta y financiada mediante el Proyecto PID TETEUME0008760TC que se desarrolla en el LICPaD (Laboratorio de Investigación en Cómputo Paralelo /Distribuido), en el cual se cuenta con la participación tanto de docentes del propio laboratorio, como de docentes del grupo LITAPS (Laboratorio de Integración de Tecnologías Aplicadas a Propotipos de Software), grupos pertenecientes al Departamento de Ingeniería en Sistemas de Información de la UTN-FRM.

1. INTRODUCCIÓN

La ciencia construye modelos, que son representaciones simplificadas de los objetos bajo estudio, con el propósito de describir, explicar o predecir alguna parte específica de la realidad. Los modelos microscópicos buscan describir el comportamiento de un ser humano al volante de un automóvil cuando es influenciado por estímulos pertenecientes a una vía de circulación. Es importante tener en cuenta que, si se describe el comportamiento de cada automóvil dentro de un sistema en todo momento, entonces podría simularse de manera casi perfecta el tráfico vehicular ya que todo lo que pase

dentro del modelo sería predecible. En general, cada agente puede representar individuos, grupos o entidades, como vehículos o peatones.

2. LÍNEAS DE INVESTIGACIÓN

La simulación de tránsito basada en agentes [1] es un enfoque específico dentro de la simulación microscópica que permite modelar cada individuo de la red (vehículos, peatones, etc.) como un agente individual con comportamientos, objetivos y reglas específicas. Luego, el sistema completo se observa desde la perspectiva de cada una de sus entidades constituyentes. Estos agentes interactúan entre sí y con su entorno, lo que permite capturar patrones emergentes y dinámicas complejas del tráfico. Existe una variedad de criterios en la comunidad científica sobre qué es un agente. No obstante, la literatura en general hace hincapié en la autonomía que presenta, lo que le permite tomar decisiones por sí mismo. Quizás la definición más simple es la de Russell [2], que considera un agente como una entidad que percibe y actúa sobre un entorno. Basándose en esta definición, se pueden caracterizar distintos tipos de agentes de acuerdo a los atributos que posean (y que van a definir su comportamiento) para resolver un determinado problema.

La característica principal de un sistema basado en agentes consiste en que los agentes son entidades discretas. En la simulación basada en agentes, un agente representa una entidad autónoma que toma decisiones basadas en un conjunto de reglas predefinidas en el modelo, su relación con el resto de los agentes, y su percepción del entorno. Los elementos clave de un modelo basado en agentes incluyen la definición de los agentes y el entorno donde se sitúan los mismos. Por otro lado, y hablando del entorno, se considera al mismo como el espacio donde los agentes interactúan, que puede ser físico (como una red vial) o abstracto (como una red social). El modelo debe poseer también reglas de interacción, que son las que determinan cómo se comunican y afectan los agentes entre sí y al entorno. Luego, de esta interacción surgen propiedades emergentes, las cuales son inesperadas, y resulta de las interacciones locales de los agentes. Además, puede incluir patrones como congestionamientos o ciclos económicos.

Los agentes cuentan con un estado interno dado por el valor de sus atributos en un momento dado. Algunos de esos atributos sirven para localizarlo espacialmente en el entorno de ejecución. Pueden percibir e interactuar tanto con otros agentes, como con el entorno a través de la percepción del mismo, y luego comportarse en base a reglas predefinidas. Algunos son capaces de aprender y adaptarse al medio. Es importante considerar que la mayoría de los sistemas basados en agentes carecen de procesos supervisores centrales, dejando la evolución del sistema en manos puramente de la interacción de los mismos. Luego, los sistemas basados en agentes pueden producir un comportamiento colectivo.

Si se dota a los agentes con suficiente cantidad de información, a través de una gran cantidad de atributos, la simulación seguramente será muy poderosa, y se podrá observar la realidad. No obstante, se enfrentan varios desafíos al incrementar el tamaño del sistema. La complejidad computacional que se afronta, supone la ejecución de múltiples procesos, y luego los recursos computacionales necesarios pueden crecer también. Por otro lado, en cuanto a la validación del modelo, resulta difícil asegurar que los resultados de la simulación puedan representar la realidad.

El desarrollo de un modelo basado en agentes involucra varias etapas. En primer lugar, se encuentra la definición del problema a resolver, donde se identifican

los objetivos del modelo y el fenómeno que se desea analizar. A continuación, se encuentra la etapa de diseño del modelo, donde se especifican las características de los agentes, sus atributos, reglas de interacción y el entorno. En la etapa de implementación se codifica el modelo utilizando plataformas destinadas a tal fin, o la utilización de algún lenguaje de programación acorde para el desarrollo del simulador. En la etapa de validación y calibración se comparan los resultados de la simulación con datos provenientes de la realidad para ajustar sus parámetros. Finalmente se encuentra la etapa de análisis, donde se realizan simulaciones para explorar diferentes escenarios y evaluar el impacto de diversas variables.

Frente al conjunto interactuante de agentes, se debe considerar también el entorno virtual donde se ubican los mismos. Este entorno, que representa al medio real en el que operan los actores reales, es el hábitat donde los agentes interactúan entre sí. Dicho entorno puede representar espacios geográficos acotados u otros ambientes, donde los agentes tienen coordenadas que indican su localización. En principio, el entorno en el que se sitúan los agentes es relativamente sencillo de diseñar y programar, por lo que en general se tiende a descuidar. Sin embargo, la influencia del entorno en el mundo real es de una importancia radical; gran parte de la complejidad que acontece en la vida de los seres humanos surge porque los individuos suelen tratar con un entorno complejo.

Para contener el entorno físico se utiliza una representación basada en un autómata celular [3]. Luego, la discretización espacial se realiza a través de la configuración de tal autómata celular, el cual cuenta con una grilla compuesta por un conjunto de celdas contiguas. Las dimensiones de la grilla son L celdas de ancho por 10 celdas de largo de 7,5 metros cada una.

En la figura 1 se puede observar la disposición de celdas transitables (marcadas con un signo -) y las celdas peatonales (marcadas con un signo #).



Fig. 1. Malla donde se encuentran celdas transitables, marcadas con un signo - y no transitables marcadas con un signo #.

Se considera un vecindario para cada celda $C_{i,j}$, compuesto este en una primera instancia por las celdas contiguas $N=\{C_{i-1,j-1}, C_{i,j-1}, C_{i+1,j-1}, C_{i-1,j}, C_{i+1,j}, C_{i-1,j+1}, C_{i,j+1}, C_{i+1,j+1}\}$, lo que representa el vecindario de Moore. El comportamiento depende del estado de cada una de sus celdas y relaciones locales, es decir, el estado de las celdas vecinas [3]. Luego, en tiempo de implementación se considera el vecindario extendido de Moore, el cual abarca celdas más alejadas.

El conjunto de estados Q al cual puede pertenecer cada celda es $Q=\{t,p\}$, donde

- t: celda transitable, y
- p: celda peatonal.

Dentro de las posibilidades que provee el modelo, se considera que el agente pueda realizar un cambio de carril. Este procedimiento consiste en buscar aquel carril en el cual el vehículo pueda alcanzar la mayor velocidad posible y comparar la nueva velocidad con la velocidad actual, procediendo a realizar el cambio de carril si la velocidad deseada es superior a la velocidad actual.
Cuando el conductor toma la decisión de realizar un cambio de carril, sigue las siguientes etapas:

- Decisión de realizar una maniobra de cambio de carril basado principalmente en la imposibilidad

de incrementar la velocidad en el propio carril,

- Seleccionar el carril donde desea situarse. (En el presente modelo solamente se cuenta con la posibilidad de cambiar al carril restante).
- Comprobación y aceptación del espacio de seguridad disponible para realizar el cambio de carril.

Se comprueba si el espacio existente entre vehículos (intervalo de seguridad) es suficiente para realizar el cambio de carril. Para ello se analiza si el espacio entre vehículos es superior a la distancia de seguridad DS y se obtiene la probabilidad de que el conductor acepte el espacio actual para realizar el cambio. En caso afirmativo se ejecuta la transición.

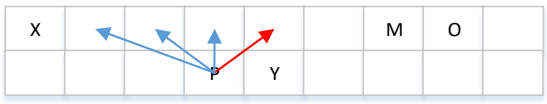


Fig 2. El agente P desea realizar un cambio de carril, con lo cual evalúa la posibilidad de sobrepaso.

En la figura 2 se puede observar que el agente P, obstruido por el agente Y, desea incrementar la velocidad en el próximo paso, con lo cual evalúa el cambio de carril (a la celda $C_{i+1,j-1}$). Para ello, el agente P tiene configurada como espacio de seguridad tres celdas: $C_{i,j-1}$, $C_{i-1,j-1}$, $C_{i-2,j-1}$.

Para el presente trabajo, se ha desarrollado una aplicación en lenguaje Java que contiene un conjunto de agentes, los cuales interactúan entre sí para simular el comportamiento de una carretera y su tráfico vehicular. En particular cada agente representa un vehículo, y es incorporado al sistema a través de un objeto proveniente de la clase de Java *Agente*. En esta clase se encapsulan los datos necesarios para asegurar su autonomía. Entre otros atributos pertenecientes a la clase *Agente* se pueden destacar:

a. N. Un nombre que permite identificar cada agente frente al resto.
b. X e Y. Valores enteros que representan una coordenada dentro de la malla.
c. A. Atributo booleano que indica si el agente prefiere velocidades mayores y tiende a realizar adelantamientos en el flujo.
d. DS. Valor entero que especifica la cantidad de celdas que un agente considera su espacio de seguridad frente al resto del flujo.
e. S. Atributo booleano que especifica si el agente ya fue sensado dentro del flujo.
f. Ca. Referencia al autómata celular sobre el cual actúa.
g. Celda. Referencia a la celda actual.
h. Vmax. Valor entero que especifica la velocidad actual.

Un agente sigue las acciones del modelo de Nagel-Schreckenberg (NASCH) [4], el cual fue propuesto en los inicios de la década de 1990. El modelo considera elementos como *aceleración* si la velocidad del vehículo es menor que su velocidad máxima y no hay vehículos delante en el flujo. En tal caso acelera en una unidad. *Desaceleración* si la distancia al vehículo de adelante es menor o igual a su velocidad actual, aquí reduce su velocidad a la distancia disponible. *Frenado aleatorio* con una probabilidad preestablecida, sucede cuando un vehículo reduce la velocidad de forma aleatoria, simulando comportamientos humanos repentinos. *Movimiento* donde los vehículos avanzan en la cuadrícula de acuerdo a la velocidad calculada. Esto permite simular el tráfico en una vía de

manera simple, pero realista. Los vehículos modifican su velocidad y avanzan en función de la distancia al vehículo de adelante y factores aleatorios. Los vehículos avanzan de una celda a otra en pasos discretos de tiempo, siguiendo reglas simples, pero que utilizadas adecuadamente permiten modelar la realidad.

La clase *Agente* hereda la clase *Thread* de Java, con lo cual cada objeto configura un hilo de ejecución independiente. La programación multihilo en Java permite que un programa ejecute múltiples tareas simultáneamente, logrando lo que se conoce como concurrencia. Los hilos permiten que diferentes partes de un programa se ejecuten simultáneamente, lo que mejora el rendimiento, especialmente en sistemas con múltiples núcleos de CPU. De esta forma se garantiza que cada proceso (agente) evolucione concurrentemente con el resto de los procesos.

Se debe observar también que diferentes procesos pueden acceder a recursos compartidos, con lo cual se torna necesario un proceso de sincronización, para que no ocurran problemas como condiciones de carrera. Se debe considerar por otro lado, que manejar la concurrencia aumenta la complejidad del código, puesto que se deben pensar cada uno de los procesos para ser ejecutados independientemente, y si los mismos comparten recursos centralizados, se pueden presentar problemas como interbloqueos o hilos en espera infinita. También se puede observar una sobrecarga de recursos, esto es, al crear demasiados hilos puede saturar la memoria y el procesador.

3. RESULTADOS OBTENIDOS

El trabajo permitió la realización de un modelo simple, basado en agentes, para la simulación del área de una carretera. La utilización de la aplicación permite observar la dinámica del modelo de Nagel-Schreckenberg, y por tanto simular situaciones de congestión. La observación de estas situaciones de congestión permite luego tomar decisiones sobre la modificación de políticas de circulación, por ejemplo tomando decisiones sobre la posible semaforización de un cruce.

4. FORMACIÓN DE RECURSOS HUMANOS

La línea es conducida principalmente por el Ing. Julio Monetti, miembro del LITAPS. Los demás participantes trabajan en diversas áreas y contribuyen en la realización de distintas líneas mediante actividades de colaboración. Del mismo modo, en el grupo se cuenta con estudiantes y becarios de grado.

5. BIBLIOGRAFÍA

[1] Gómez-Cruz, N. Simulación basada en agentes: Una metodología para el estudio de sistemas complejos. DOI: 10.2307/j.ctvfc5506.12. 2018

[2] Russell, S: Inteligencia Artificial: un enfoque moderno. Prentice - Hall. México, 1996.

[3] Vargas-Forero, V, Muñoz Ceballos, L, García-Baños, A, Jaramillo-Molina, C. Modelo con autómatas celulares para analizar la accesibilidad peatonal al interior del campus universitario Meléndez de la Universidad del Valle. Scientia Et Technica, vol. 24, núm. 1, pp. 67-75, 2019

[4] Schadschneider, A. The Nagel-Schreckenberg model revisited. arXiv:cond-mat/9902170v1. 2018.

Sistema para análisis y sustitución de abreviaturas en dominios médicos.

Marcelo G. Muzas and Sandra E. Roger

marcelo.muzas@est.fi.uncoma.edu.ar
roger@fi.uncoma.edu.ar

Grupo de Investigación en Lenguajes e Inteligencia Artificial
Departamento de Teoría de la Computación - Facultad de Informática
UNIVERSIDAD NACIONAL DEL COMAHUE

Resumen

En el ámbito de la salud, la transición hacia registros médicos digitales ha superado las deficiencias inherentes a los formatos en papel, facilitando un acceso expedito a la información del paciente. No obstante, se postula que el valor intrínseco de estos registros reside en su potencial para un análisis inteligente, susceptible de mejorar la calidad de la atención médica, la realización de estudios estadísticos robustos y el soporte a la toma de decisiones clínicas informadas. Un obstáculo significativo para el procesamiento eficiente de esta información médica radica en la carencia de normalización en los textos, caracterizada por la presencia de errores tipográficos y ortográficos, así como el uso extendido de abreviaturas no estandarizadas. Esta heterogeneidad dificulta el análisis automatizado, limitando la explotación del potencial inherente a los datos clínicos disponibles. En respuesta a este desafío, el presente trabajo propone el desarrollo de una herramienta basada en técnicas de aprendizaje profundo (*deep learning*) para la identificación y normalización de abreviaturas en textos médicos. El objetivo general de esta investigación es optimizar las evoluciones digitales de los pacientes mediante la detección y reemplazo de abreviaturas, enriqueciendo su contexto para su posterior integración en sistemas de registros electrónicos y análisis de datos clínicos.

Palabras Clave: Generación de Lenguaje Natural, Aprendizaje Automático, Código, .

Contexto

Este trabajo está parcialmente financiado por la UNCo, en el marco del nuevo proyecto de investigación *Tecnologías Semánticas para el desarrollo de Agentes Inteligentes*. El proyecto de investigación tiene una duración de cuatro años y ha comenzado en 2022.

1. Introducción

En el contexto actual de la atención médica, la digitalización de la información clínica ha experimentado una adopción generalizada. Los profesionales de la salud registran de manera electrónica tanto los diagnósticos como las "evoluciones" clínicas de los pacientes, entendiendo por evolución el registro detallado del progreso, tratamiento y estado general de un paciente a lo largo del tiempo. Esta transición hacia los registros digitales ha mitigado significativamente las problemáticas inherentes a los registros en papel, tales como la dificultad en la interpretación de diversas caligrafías, la necesidad de un espacio físico considerable para el almacenamiento de documentos y la limitada agilidad en la recuperación de la información. Los sistemas digitales permiten un acceso rápido y eficiente

a la historia clínica de un paciente utilizando datos básicos como el Documento Nacional de Identidad (DNI) y la fecha de atención.

Sin embargo, el verdadero valor añadido de estos sistemas digitales reside en la capacidad de realizar un procesamiento inteligente de los datos que contienen. Al someter estos textos a un análisis avanzado, se abre la posibilidad de obtener información valiosa para mejorar la calidad de la atención médica, facilitar la realización de estudios estadísticos exhaustivos y proporcionar un soporte más sólido a la toma de decisiones clínicas. Un desafío clave en este procesamiento es la falta de normalización en los textos médicos. La presencia de errores de tipeo y ortográficos, así como el uso extensivo de abreviaturas, particularmente aquellas pertenecientes al dominio médico, son comunes. Muchas de estas abreviaturas no son reconocidas por las herramientas tradicionales de análisis de texto, lo que dificulta su procesamiento automatizado y limita su utilidad potencial.

El desarrollo del plan de trabajo se realizará en el marco del proyecto de investigación "Tecnologías Semánticas para el desarrollo de Agentes Inteligentes". En dicho proyecto de investigación se desarrolla una línea de investigación que explora sobre temas afines tanto al análisis y desarrollo de técnicas tanto de aprendizaje profundo como del Procesamiento en Lenguaje Natural con el objetivo de dar soporte a los agentes inteligentes. Particularmente, se ha escogido experimentar sobre herramientas innovadora basadas en *deep learning* que den soporte en la identificación y normalización de textos médicos, mejorando así la calidad y eficiencia del procesamiento automatizado de datos clínicos. Esto facilitará su uso en análisis avanzados, toma de decisiones y estudios estadísticos en el ámbito de la salud.

2. Línea de Investigación y Desarrollo

El proyecto de investigación *Tecnologías Semánticas para el Desarrollo de Agentes Inteligentes*. El objetivo general de este proyecto es generar conocimiento especializado en el ámbito de los agentes inteligentes capaces de acceder, procesar y recuperar información mediante el uso de tecnologías semánticas. Esta línea de investigación busca potenciar la capacidad de los sistemas para interpretar y extraer significado de los datos de manera más efectiva, avanzando hacia una comprensión automatizada y contextualizada de la información.

Dentro de este marco, se ha identificado una línea de investigación específica centrada en el desarrollo e implementación de una herramienta innovadora para la identificación y normalización de abreviaturas en textos médicos. Este enfoque no solo pretende mejorar la calidad y eficiencia del procesamiento automatizado de datos clínicos, sino que también se sustenta en la aplicación de técnicas avanzadas de *deep learning* [1, 2, 3]. Al abordar el desafío de las abreviaturas en los registros clínicos, se espera contribuir significativamente a la optimización del análisis de información en el ámbito de la salud, facilitando la generación de datos más precisos y contextualizados que respalden la toma de decisiones clínicas y mejoren la atención al paciente.

Las abreviaturas son formas abreviadas de cadenas de texto las cuales son ampliamente utilizadas en la literatura biomédica, notas clínicas, artículos científicos, entre otras. Estas abreviaturas pueden ser ambiguas, especialmente en textos breves o profesionales donde no se proporcionan sus definiciones. Por ejemplo, aproximadamente el 15 % de las consultas en PubMed incluyen abreviaturas [4] , y alrededor del 14,8 % de todos los tokens en un conjunto de datos de notas clínicas son abreviaturas [5] . En ambos casos, rara vez se incluyen las definiciones de estas abreviaturas, por lo que la expansión automática de abreviaturas ambiguas a sus formas completas es fundamental en los sistemas de procesamiento de lenguaje natural (PLN) aplicados al ámbito biomédico.

Para alcanzar este objetivo general, se han establecido una serie de objetivos específicos. En primer lugar, se desarrollará un diccionario exhaustivo de abreviaturas y términos médicos utilizados en las evoluciones clínicas digitales de los pacientes. Este diccionario servirá como base para el entrenamiento de un sistema de de-

tección y reemplazo de abreviaturas. En segundo lugar, se diseñará e implementará un sistema basado en expresiones regulares para identificar y normalizar abreviaturas en las evoluciones clínicas, evaluando su efectividad y precisión en un entorno clínico real.

Adicionalmente, se construirá un corpus anotado que incluya una amplia variedad de evoluciones clínicas digitales, con ejemplos de abreviaturas y su contexto. Este corpus proporcionará datos representativos y etiquetados necesarios para el entrenamiento de un modelo de *deep learning*, asegurando una cobertura adecuada de diversas especialidades médicas y escenarios clínicos. Finalmente, se llevará a cabo una evaluación comparativa de la efectividad y eficiencia de ambos sistemas (el basado en expresiones regulares y el basado en deep learning), analizando métricas como precisión, tiempo de procesamiento y capacidad de contextualización de la información. Este análisis permitirá determinar qué metodología ofrece mejores resultados en el enriquecimiento de las evoluciones clínicas digitales.

Se espera que esta propuesta contribuya de manera significativa a la mejora de la calidad y eficiencia del procesamiento de textos médicos, sentando las bases para un análisis automatizado más preciso y útil en el ámbito clínico. Asimismo, este trabajo aspira a establecer un precedente en la aplicación de tecnologías semánticas y de aprendizaje profundo para la normalización de datos médicos, promoviendo avances en la interoperabilidad y el uso efectivo de la información en salud.

3. Resultados Obtenidos y Trabajos Futuros

En este trabajo, se implementó un prototipo de herramienta para la detección y desambiguación de abreviaturas en textos en español dentro del dominio médico. La implementación se estructuró en dos etapas principales.

En la primera fase, se diseñará un detector de abreviaturas basado en expresiones regulares, capaz de identificar patrones comunes como palabras terminadas en punto, secuencias de caracteres en mayúscula y otras estructuras típicas de abreviaturas. Esta etapa permitirá generar un conjunto inicial de posibles abreviaturas presentes en los textos analizados.

La segunda fase abordará el problema de la desambiguación de abreviaturas mediante un proceso de *Word Sense Disambiguation* (WSD). Dado que una misma abreviatura (*Short Form*, SF) puede tener múltiples formas extendidas (*Long Form*, LF) según el contexto, se implementará un mecanismo para inferir la interpretación más adecuada en cada caso, considerando variables semánticas y contextuales.

Adicionalmente, se incorporará un enfoque basado en *deep learning* para mejorar la precisión de la primera fase del sistema. Se realizará un *fine-tuning* del modelo *es_core_news_md*, que forma parte de la familia de modelos de procesamiento de lenguaje natural de spaCy[1]. Este modelo, disponible en diferentes tamaños (sm, md y lg), será ajustado utilizando un corpus etiquetado con textos que contienen abreviaturas, con el objetivo de optimizar la tarea de detección de abreviaturas.

Como trabajo futuro, se plantea el desarrollo del modelo de desambiguación mediante técnicas más avanzadas de deep learning, así como la ampliación del corpus de entrenamiento. También se explorará la integración de la herramienta en flujos de procesamiento de texto en tiempo real y su aplicación en el análisis de documentos médicos ya redactados

4. Formación de Recursos Humanos

Durante la realización de esta investigación se espera lograr, la culminación de 1 (posiblemente 2) tesis de grado dirigidas y/o codirigidas por los integrantes del proyecto.

Así también, se espera que durante el desarrollo del proyecto, los integrantes del mismo estén en proceso de elaboración de su tesis de posgrado puedan consolidar su formación en in-

[1] spacy.io

vestigación, y que el trabajo realizado contribuya a su graduación.

Finalmente, es constante la búsqueda hacia la consolidación como investigadores de los miembros más recientes del grupo.

Referencias

[1] Ismat Mohd Sulaiman, Awang Bulgiba, Sameem Abdul Kareem, and Abdul Aziz Latip. Deciphering abbreviations in malaysian clinical notes using machine learning. *Methods of Information in Medicine*, 2025.

[2] Mucahit Cevik, Sanaz Mohammad Jafari, Mitchell Myers, and Savas Yildirim. Sequence labeling for disambiguating medical abbreviations. *Journal of Healthcare Informatics Research*, 7(4):501–526, 2023.

[3] Areej Jaber and Paloma Martínez. Disambiguating clinical abbreviations using pretrained word embeddings. In *HEALTHINF*, pages 501–508, 2021.

[4] Rezarta Islamaj Dogan, G Craig Murray, Aurélie Névéol, and Zhiyong Lu. Understanding pubmed® user search behavior through log analysis. *Database*, 2009:bap018, 2009.

[5] Hua Xu, Peter D Stetson, and Carol Friedman. A study of abbreviations in clinical notes. In *AMIA annual symposium proceedings*, volume 2007, page 821, 2007.

Traducción de Lengua de Señas con Deep Learning

Facundo Quiroga[1,2], Franco Ronchetti[1,2], Gastón Rios[1,3], Pedro Dal Bianco[1,3], Santiago Ponte Ahon[1,3], Oscar Stanchi[1,4], Juan Seery[1], Laura Lanzarini[1], Waldo Hasperué[1]

[1] Instituto de Investigación en Informática LIDI, Facultad de Informática, Universidad Nacional de La Plata, La Plata, Argentina.*
[2] Comisión de Investigaciones Científicas de la Pcia. De Bs. As. (CICPBA)
[3] Becario postgrado UNLP
[4] Becario CONICET
*** Centro asociado de la Comisión de Investigaciones Científicas de la Pcia. De Bs. As. (CIC)**

Contacto: fquiroga@lidi.info.unlp.edu.ar

CONTEXTO

Esta presentación corresponde a algunas de las tareas de investigación que se llevan a cabo en el III-LIDI en el marco del proyecto "Inteligencia de Datos. Técnicas y Modelos de Machine Learning" perteneciente al Programa de Incentivos (2023-2026).

RESUMEN

Esta línea de investigación se centra en el estudio y desarrollo de sistemas y modelos de Deep Learning para la Traducción de Lengua de Señas (TLS), con Redes Neuronales Convolucionales (CNNs), Recurrentes (RNNs) y Transformers. En particular, se trabaja en:

- Estudio y diseño de Modelos de Traducción de la Lengua de Señas
- Desarrollo de sistemas de público para la traducción de la Lengua de Señas Argentina (LSA).
- Generación, selección y preprocesamiento de Bases de Datos de Lengua de Señas.
- Generación de ejemplos sintéticos de LSA para paliar la falta de datos.
- Plataforma de crowdsourcing con videos de la LSA.

Palabras clave: Redes Neuronales, Redes Convolucionales, Redes Recurrentes, Transformers, Visión por Computadoras, Lengua de Señas, Bases de datos, *Crowdsourcing*, Redes Generativas Adversarias

1. INTRODUCCION

El Instituto de Investigación en Informática LIDI (III-LIDI) tiene una larga trayectoria en el estudio, investigación y desarrollo de distintos métodos de Aprendizaje Automático y Redes Neuronales.

En particular, lleva diez años trabajando en esta línea de investigación aplicada a la Traducción de Lengua de Señas (TLS). TLS busca traducir entre un video de lengua de señas y un texto correspondiente en una lengua escrita. Alternativamente, el Reconocimiento de Lengua de Señas (RLS) es una tarea intermedia que busca traducir entre un video de lengua de señas y las etiquetas de cada seña, llamadas glosas.

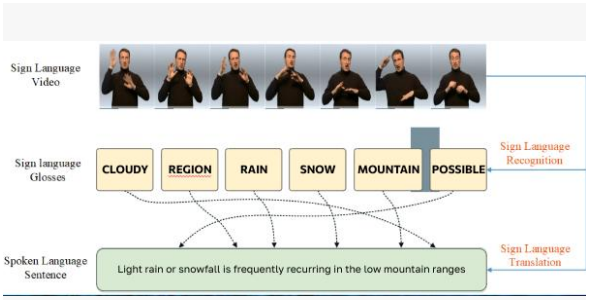


Figura 1. Ejemplo de TLS utilizando glosas. En un primer paso, se realiza un reconocimiento de las señas de forma individual. Luego, se traduce de las glosas de lengua de señas a la lengua escrita, como el inglés.

TLS es un campo de estudio activo que abarca la interacción humano-computadora, la visión por computadora, el procesamiento de lenguaje natural y el aprendizaje automático. Es un problema complejo y multidisciplinar, que presenta diversos subproblemas a resolver como el reconocimiento del intérprete que realiza una seña, la segmentación de manos, la clasificación de diferentes configuraciones y de un gesto dinámico, entre otros.

La falta de datos es la dificultad principal para lograr modelos de TLS de alto desempeño. Por eso, además de desarrollar nuevos modelos de TLS, exploramos tres vías distintas para paliar dicha falencia: creación de datos sintéticos con modelos generativos, crowdsourcing para aumentar el volumen y mejorar la calidad de LSA-T, y una librería de carga y procesamiento unificado de bases de datos de lenguas de señas para realizar transferencia de aprendizaje.

1.1. LSA-T: Base de datos de LSA para TLS

La creación de un conjunto de datos apropiado para la LSA resulta esencial para la creación de modelos específicos. Por este motivo, se continúa con el desarrollo y la validación de la base de datos LSA-T [1], la más grande al momento con señas de la LSA y una de las más grandes del mundo en general.



Figura 2 . Ejemplo de video de la bases de datos LSA-T.

Además, se puso en marcha la plataforma **seni.ar** para permitir que la comunidad sorda contribuya con nuevos videos de LSA, y agregue información sobre los existentes.

1.2. Generación de videos de Lengua de Señas basados en poses.

Se desarrollaron modelos basados en Redes Generativas Adversarias (GANs) y modelos Transformer para la generación de señas individuales en formato de pose, con el objetivo de mejorar el entrenamiento de modelos de RLS. Los modelos obtenidos permiten condicionar al generador utilizando el inicio de una seña, y de esta forma generar videos de lengua de seña arbitrarios.

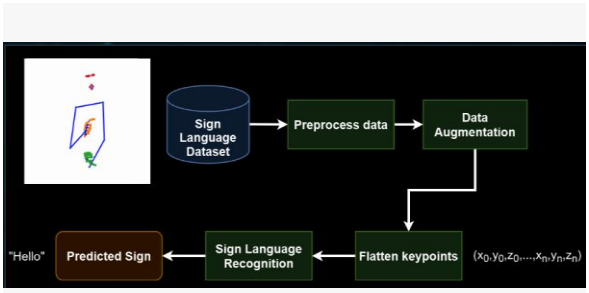


Figura 3. Entrenamiento con generación de datos sintéticos

En particular, se encontró que pre-entrenar los modelos de TLS con datos generados sintéticamente mejora su desempeño, en comparación con no preentrenar, preentrenar con otros datos, o utilizar otros tipos de aumentación de datos [3].

Además, se propuso el modelo ConvAtt, basados en poses, que es significativamente más ligero que otros y tiene un desempeño comparable con el estado del arte [3].

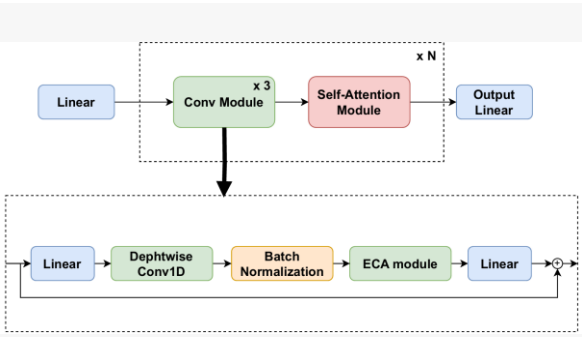


Figura 4 . Modelo ConvAtt para la generación de muestras sintéticas.

1.3. Modelos de Traducción de Lengua de Señas basados en Poses y sin glosas

Para promover un uso amplio de los modelos de TLS, estos deben ser ligeros, y además funcionar en diversas condiciones de

luminosidad, fondos, etc. Además, dada la falta de datos del área, y en particular datos etiquetados con las glosas, es interesante que los modelos puedan aprovechar todo tipo de información disponible. Por ese motivo, proponemos utilizar modelos de TLS que utilizan poses como entrada y no emplean la predicción de glosas como tarea intermedia. Para ello, se desarrolló un modelo basado en Transformers de tipo Encoder/Decoder. Se encontró que si bien el desempeño de este tipo de modelos es inferior al de otros que utilizan videos y/o glosas, su capacidad de traducción es significativa, manteniendo una baja complejidad computacional, lo cual valida este enfoque como un posible camino a seguir [4].

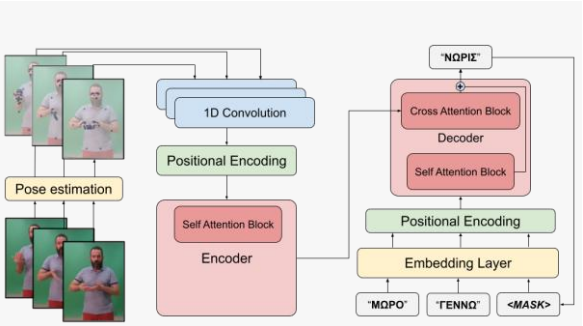


Figura 5. Modelo propuesto para traducción basada en Poses

Dataset	Method	Accuracy	BLEU-4	BLEU-3	BLEU-2	BLEU-1
LSA-T	Greedy	16.7%	0.2	0.3	5	6.4
LSA-T	Beam	16.7%	0.05	0.1	5	6.7
RWTH	Greedy	41.7%	5.7	6.85	9.98	15.87
RWTH	Beam	41.7%	5.9	6.85	10	18.87
GSL	Greedy	93.4%	43.06	54.55	63.45	75.46
GSL	Beam	93.4%	43.74	55.15	63.2	75.78

Figura6.Resultados del modelo en 3 grandes bases de datos de TLS: LSA-T, RWTH y GSL

1.4. Librería de carga y preprocesamiento de Bases de Datos de Lengua de Señas

Una de las trabas más frecuentes al experimentar con varias bases de datos de SLT es la diversidad en sus formatos, tamaños, y modo de acceso. Además, en general las estrategias de preprocesamiento utilizadas en los experimentos actuales suelen diferir. Para facilitar las comparaciones y bajar la barrera de entrada al área, se desarrollaron las librerías SLTD[1], que permite descargar y cargar fácilmente diversas bases de SLT y Posecraft[2], una librería para preprocesar datos de poses, como las usadas en TLS. En ambos casos, las librerías son de código abierto, y están orientadas a usarse en el contexto de entrenar y evaluar modelos del framework PyTorch.

```
from slt_datasets.SLTDataset import SLTDataset
dataset = SLTDataset(
    data_dir="/path/to/dataset",
    input_mode="video",
    output_mode="text",
    split="train"
)
```

Figura 7. Uso de la librería **SLTD** para la descarga y carga de bases de datos de TLS.

```
# Load a pose from a .npy file
pose_data = Pose(path="path/to/pose.npy")

# Create a set of transforms
transforms = torch.nn.Sequential(
    CenterToKeypoint(center_keypoint=0),
    NormalizeDistances(indices=(11, 12),
distance_factor=0.2))
# Apply transforms
transformed_pose = transforms(pose_data.pose)
```

Figura 8. Uso de la librería **posecraft** para el procesamiento de

[1] https://github.com/pedroodb/slt_datasets

[2] https://github.com/pedroodb/posecraft

2. LINEAS DE INVESTIGACIÓN Y DESARROLLO

- Redes neuronales profundas, convolucionales y transformers.
- Traducción de Lengua de Señas
- Modelos ligeros de TLS basados en poses y sin glosas.
- Generación de videos con Redes Neuronales para lidiar con la falta de datos y desbalance de clases en Lengua de Señas
- Herramientas para la manipulación de Bases de Datos de Lengua de Señas.

3. RESULTADOS OBTENIDOS/ESPERADOS

- Validación de un modelo generador de videos de lengua de señas basados en poses, generalizable a cualquier conjunto de datos y poses.
- Ampliación de LSA-T mediante sistema web de Crowdsourcing.
- Diseño de modelos ligeros basados en poses y sin glosas para la TLS
- Librerías para la carga y manipulación de bases de datos de TLS basadas en poses.

4. FORMACIÓN DE RECURSOS HUMANOS

El grupo de trabajo de la línea de I/D aquí presentada está formado por: 3 profesores con dedicación exclusiva, un JTP dedicación exclusiva, 2 investigadores CIC-PBA, 2 becarios de posgrado de la UNLP con dedicación docente. Actualmente se están desarrollando 2 tesis de doctorado y 1 de maestría en este tema. También participan en el desarrollo de las tareas becarios y pasantes del III-LIDI.

5. REFERENCIAS

[1] Dal Bianco, P., Ríos, G., Ronchetti, F., Quiroga, F., Stanchi, O., Hasperué, W., & Rosete, A. (2023, January). LSA-T: The First Continuous Argentinian Sign Language Dataset for Sign Language Translation. In Advances in Artificial Intelligence–IBERAMIA 2022: 17th Ibero-American Conference on AI, Cartagena de Indias, Colombia, November 23–25, 2022, Proceedings (pp. 293-304). Cham: Springer International Publishing.

[2] Ríos, G., Dal Bianco, P., Ronchetti, F., Quiroga, F., Stanchi, O., Hasperué, W. Generación de gestos de lengua de señas con redes neuronales generativas basadas en poses y etiquetas. CACIC 2023, Junin, Buenos Aires

[3] Ríos, G.G., Dal Bianco, P.A., Ronchetti, F., Quiroga, F.M., Ponte Ahón, S.A., Stanchi, O.A. and Hasperué, W., 2024. Scaling up ConvAtt for Sign Language Recognition. In XXX Congreso Argentino de Ciencias de la Computación (CACIC)(La Plata, 7 al 11 de octubre de 2024).

[4] Dal Bianco, P.A., Ríos, G.G., Hasperué, W., Stanchi, O.A., Ronchetti, F. and Quiroga, F.M., 2024. Gloss-free Argentinian Sign Language Translation with pose-based deep learning models. In XXX Congreso Argentino de Ciencias de la Computación (CACIC)(La Plata, 7 al 11 de octubre de 2024).

UNA GRAMÁTICA PARA UNA LENGUA EN SITUACIÓN DE RESISTENCIA: EL CASO DEL MAPUCHE

Leonardo A. Correa[1] Veronica Dahl[2] Jorge P. Rodriguez[1]
Laura A. Cecchi[1] Lucas Curapil[3]

email: leonardo.correa@est.fi.uncoma.edu.ar, veronica_dahl@sfu.ca,
{j.rodrig, lcecchi}@fi.uncoma.edu.ar, curapil@yahoo.com.ar

[1] *Grupo de Investigación en Lenguajes e Inteligencia Artificial*
Facultad de Informática
UNIVERSIDAD NACIONAL DEL COMAHUE
Neuquén - Argentina

[2] School of Computing Science
SIMON FRASER UNIVERSITY
Burnaby, B.C. - Canadá

[3] Departamento de Lenguas Originarias
Facultad de Lenguas
UNIVERSIDAD NACIONAL DEL COMAHUE
Neuquén - Argentina

RESUMEN

El pueblo Mapuche es uno de los más importantes en el sur de Argentina y Chile, pero su lengua se encuentra en situación de vulnerabilidad.

Las Womb Grammars son una posibilidad para la inferencia gramatical de lenguas poco estudiadas, buscando simplificar la tarea de los lingüistas de todo el mundo.

Partiendo desde una lengua fuente bien estudiada y mediante el uso del modelo probado exitosamente por Veronica Dahl, esta línea de investigación y desarrollo propone inferir una gramática para un subconjunto de frases nominales de la lengua Mapuche, contribuyendo de este modo a su preservación.

Palabras Clave: Gramáticas basadas en restricciones - Womb Grammars - Lengua Mapuche

CONTEXTO

Esta línea de investigación se desarrolla en el contexto de los temas de interés que promueve el Grupo de Investigación en Lenguajes e Inteligencia Artificial (GILIA), de la Facultad de Informática. En particular, se enmarca en el ámbito de dos proyectos de investigación del GILIA, ambos financiados por la Universidad Nacional del Comahue: Modelos formales, agentes inteligentes y aplicaciones para la enseñanza de las Ciencias de la Computación (04/F022) y Tecnologías semánticas para el desarrollo de agentes inteligentes (04/F020).

Este trabajo está financiado por la Universidad Nacional del Comahue, en el marco de los proyectos antes mencionados. Los proyectos de investigación tienen una duración de cuatro años, a partir de enero de 2022.

Además, el trabajo se realiza en colaboración con el grupo de investigación de Verónica Dahl, de la Simon Fraser University, financiado por NSERC de Canadá (NSERC grant 31611021).

Asimismo, se trabaja en colaboración con la Cátedra Libre y Anual de Idioma y Cultura Mapuche, perteneciente al Departamento de Lenguas Originarias, de la Facultad de Lenguas, Universidad Nacional del Comahue.

1. INTRODUCCIÓN

La región de la Patagonia argentina y la Araucanía chilena son el hogar de numerosos pueblos originarios con diferentes estilos de vida, creencias e idiomas. Entre los más prominentes se encuentra el pueblo Mapuche, uno de los más numerosos de la región. En la Argentina, este pueblo se ha reducido a una cantidad pequeña de comunidades y los hablantes de la lengua Mapuche prácticamente no existen. Por esta razón, se considera que su cultura e idioma están en situación de vulnerabilidad [1,2]. Actualmente, se observa un fuerte proceso de revitalización llevado a cabo desde organizaciones, comunidades, grupos y familias.

La lengua es uno de los pilares más poderosos de una cultura y uno de los derechos humanos[2], y es por esto que su estudio y preservación resultan esenciales. En esta dirección, la Universidad Nacional del Comahue promueve el desarrollo de una perspectiva intercultural en la educación superior, impulsando el estudio y difusión de la lengua Mapuche a través de diferentes iniciativas expresadas en su estatuto.

El Mapuche es una lengua tradicionalmente oral y poco estudiada formalmente, y que no tiene una escritura universal y estandarizada. En la actualidad, no existe un acuerdo entre los lingüistas que determine la manera de formular formalmente las reglas del lenguaje hablado y la correcta escritura de la lengua.

Las gramáticas son importantes en el estudio, la formalización y preservación de cualquier lenguaje, ya sea humano o de programación. Las gramáticas de Chomsky, particularmente, han sido fundamentales para las Ciencias de la Computación, siendo usadas extensamente para especificar lenguajes de programación[3]. Estas gramáticas poseen reglas de generación del estilo “S ← A”, por ejemplo, “Oración ← Sujeto, Verbo”.

Estas reglas resultan ideales para caracterizar y formalizar, evitando ambigüedades, a los lenguajes de programación.

Sin embargo, el lenguaje humano es mucho más flexible. Si representásemos oraciones mediante la regla antes propuesta, entonces un cambio en un lenguaje que utiliza una estructura “sujeto, verbo, objeto” requeriría cambiar una gran cantidad de reglas que son utilizadas para formar distintos tipos de oraciones. Por esto, resulta más adecuado el uso de gramáticas basadas en satisfacción de restricciones. Las gramáticas de restricciones no están basadas en la estructura de una frase, sino en las propiedades que ésta cumple, entonces simplemente podríamos cambiar la propiedad que indica el orden en que el sujeto, objeto y verbo aparecen en una oración, cambiando así todas las reglas que usen esa propiedad. Mediante estas gramáticas ya se han desarrollado métodos que permiten inferir reglas de lenguajes humanos[4,5,6].

En esta línea de investigación y desarrollo se propone avanzar en la definición de una gramática para la lengua Mapuche que pueda ser automatizada.

Se espera que esto contribuya significativamente a la preservación de la lengua Mapuche, mediante su difusión y la facilitación de su enseñanza.

Este trabajo está estructurado como sigue. En la siguiente sección, se describe la línea de investigación y desarrollo. En la Sección 3, se indican algunos resultados obtenidos y esperados. Finalmente, en la última sección se comentan aspectos referentes a la formación de recursos humanos en esta temática.

2. LÍNEA DE INVESTIGACIÓN Y DESARROLLO

A pesar de que existen escritos sobre la lengua Mapuche y sus reglas gramaticales [7, 8], no hay un estándar aceptado y muchos elementos de la escritura actualmente se encuentran en discusión y en evolución constante. Esta situación se debe, entre otras razones, a que el territorio Mapuche es muy amplio, existiendo más de ocho variedades de la lengua, adoptando usos diversos de las acentuaciones y modificaciones de consonantes, entre otras variaciones.

En la literatura se pueden encontrar desarrollos que involucran la construcción de un modelo capaz de inferir una gramática para las frases nominales de un idioma desconocido a partir de otro conocido, mediante el uso de gramáticas de satisfacción de restricciones llamadas Womb Grammars[4]. Este modelo resultó apropiado para inferir una gramática para las frases nominales del Yorùbá, una lengua Nigeriana, y del Ch'ol, una lengua Mexicana. Ambas han sido estudiadas y, por lo tanto, resultan buenos casos de prueba.

En vista de lo anterior, en esta línea de investigación se propone desarrollar una gramática basada en restricciones que nos permita inferir una gramática circunscrita a un subconjunto de frases nominales de la lengua Mapuche. En este sentido, nos basaremos en el método y modelo que ya ha sido probado exitosamente por Verónica Dahl con las lenguas Yorùbá y Ch'ol [5,6].

Este trabajo se desarrolla en el marco de las actividades de los proyectos de investigación *Tecnologías Semánticas para el desarrollo de agentes inteligentes y Modelos Formales, Agentes Inteligentes y Aplicaciones para la Enseñanza de las Ciencias de la Computación*. Por un lado, la generación de gramáticas que formalicen lenguajes naturales son parte de los objetivos del proyecto de investigación *Tecnologías Semánticas para el desarrollo de Agentes Inteligentes*.

Por otra parte, el proyecto *Modelos Formales, Agentes Inteligentes y Aplicaciones para la Enseñanza de las Ciencias de la Computación*, busca desarrollar construcciones teóricas que contribuyan a la caracterización de modelos formales y la definición de sistemas inteligentes orientados al procesamiento de información. En este sentido, se plantea como objetivo el diseño, implementación y evaluación de herramientas que los soporten.

3. RESULTADOS OBTENIDOS Y ESPERADOS

Inicialmente, establecimos la estructura organizacional de base que soporta el desarrollo de la línea de investigación. En este sentido, se contactaron lingüistas especialistas en la lengua Mapuche, como así también a grupos de investigación que han trabajado en gramáticas de lenguas en situación de resistencia.

Analizando el estado del arte sobre inferencia de gramáticas, encontramos trabajos basados en Womb Grammars[4,6,7] y en métodos no supervisados, que además testean LLM (Large Language Model)[9]. En esta primera etapa se utilizará como técnica a las Womb Grammars, porque permiten inferir una gramática para una lengua a partir de otra diferente, lo que podría simplificar enormemente el trabajo de lingüistas que buscan conservar lenguas en estado de vulnerabilidad. En este sentido, las Womb Grammars proveen una metodología que basa la inferencia a partir de las propiedades del idioma. Sobre estas propiedades se generan restricciones, considerando una lengua sintácticamente conocida y un conjunto de frases correctamente escritas en la lengua objetivo, en nuestro caso, el Mapuche. Existen al menos 3 lenguas para las que ya se han desarrollado

gramáticas y que pueden ser utilizadas como idiomas fuentes: Inglés, Ch'ol y Yorùbá [6,7]. Actualmente, nos encontramos, por una parte, analizando las frases nominales del idioma Mapuche, las cuales conformarán el corpus de entrada a un Womb Grammar Parser. A partir de este corpus, el Parser nos indicará las propiedades sintácticas que fallan, cuyo estudio y modificación llevará a inferir la gramática buscada.

Por otro lado, estamos estudiando las características de los tres idiomas que pueden ser utilizados como lengua fuente. Es nuestra idea, seleccionar aquel que tenga la mayor similitud con el Mapuche, con el propósito de reducir la complejidad al inferir la nueva gramática.

Para la implementación de dicha gramática utilizaremos el lenguaje Prolog, que es bien conocido por su poder para el procesamiento de lenguaje natural.

Esperamos que esta gramática sea capaz de reconocer la correctitud de un subconjunto de frases nominales en Mapuche y resaltar los errores en el caso de frases incorrectas.

4. FORMACIÓN DE RECURSOS HUMANOS

El primer autor está desarrollando su tesis de Licenciatura en Ciencias de la Computación en el tema de esta línea.

Por otra parte, otro de los autores de este trabajo está desarrollando su trabajo final integrador de la Maestría en Enseñanza en Escenarios Digitales, posgrado que dicta la AUSA (Asociación de Universidades Sur Andina), de la que la Universidad Nacional del Comahue forma parte.

El último de los autores está inscripto en el Doctorado en Antropología, de la UBA y actualmente está en etapa de redacción de su tesis doctoral.

5. BIBLIOGRAFÍA

[1]Pfefferle, Jana. "The Chilean Interest in the Conservation and Learning of the Native Language Mapudungun in Chile." (2015).

[2]Jerez, C. El mapuzugun, una lengua en situación de resistencia https://news.un.org/es/story/2019/04/1454571. Noticias ONU. Mirada global Historias humanas.

[3]Qureshi, Muhammad Aasim, Muhammad Asif, and Saira Anwar. "NewBee: Context-Free Grammar (CFG) of a New Programming Language for Novice Programmers." *INTELLIGENT AUTOMATION AND SOFT COMPUTING* 37.1 (2023): 439-453.

[4]Dahl, Veronica, and J. Emilio Miralles. "Womb grammars: Constraint solving for grammar induction." *Proceedings of the 9th Workshop on Constraint Handling Rules. vol. Technical report CW*. Vol. 624. 2012.

[5]Adebara, Ife, and Veronica Dahl. "Grammar Induction as Automated Transformation between Constraint Solving Models of Language." *KnowProS@ IJCAI* 7 (2016).

[6]Dahl, Veronica, et al. "Grammar induction for under-resourced languages: the case of ch'ol." *Analysis, Verification and Transformation for Declarative Programming and Intelligent Systems: Essays Dedicated to Manuel Hermenegildo on the Occasion of His 60th Birthday*. Cham: Springer Nature Switzerland, 2023. 113-132.

[7]Hernández Sallés, Arturo, Nelly Ramos Pizarro, and Rosa Huenchulaf Cayuqueo. "Gramática básica de la Lengua Mapuche." (2011).

[8]Tripailaf, Elisa, Nicacio Carmelo

Antinao, and Lucas Curapil. "Aprenda idioma mapuche: quim mapuche zungu ngue 1." (2010).

[9]da Silva, Diego Pedro Gonçalves, and Thiago Alexandre Salgueiro Pardo. "Grammar Induction for Brazilian Indigenous Languages." *Proceedings of the 16th International Conference on Computational Processing of Portuguese-Vol. 2*. 2024.

Uso de Inteligencia Artificial como asistente del Proceso de Desarrollo de Software – un caso testigo

Luis Romero[1], Carlos Gustavo Lopez[1], Enrique Fernandez[1], María Alejandra Ochoa[1], Eugenia Artola[2]

Universidad Nacional del Oeste[1]

Universidad Nacional de Cuyo[2]

gustavolopez/enfernandez/mochoa /@uno.edu.ar, ecartola@hotmail.com, romeroluis.uno@gmail.com

Resumen

El uso de la inteligencia artificial (IA) en el desarrollo de software ha revolucionado la forma en que se gestionan los proyectos informáticos. A través de herramientas como el ChatGPT, se han optimizado tareas en distintas fases del ciclo de vida del software, desde la definición de requerimientos hasta el testing. Este artículo resume una experiencia aplicada en un proyecto de software, destacando cómo la IA puede asistir en la documentación, análisis y automatización de tareas, reduciendo tiempos y mejorando la calidad del proceso.
El desarrollo del software fue realizado por alumnos del último año de la Licenciatura en Informática de la Universidad Nacional de Oeste bajo la tutela de docentes del grupo de investigación del Instituto de Ingeniería y Nuevas Tecnologías de dicha Universidad.

Palabras clave: *Inteligencia Artificial (IA), Desarrollo de Software, Automatización, Casos de Uso, Testing de Software, Generación de Código, Optimización de Procesos.*

Contexto

El proyecto "Gestión ágil y unificada basada en mediciones de software" que lleva adelante el Instituto de Ingeniería y Nuevas Tecnologías de la Universidad Nacional del Oeste, tiene como objetivo, entre otros, definir nuevos indicadores que permitan hacer una gestión más eficiente de los proyectos de desarrollo de software en un contexto cada vez más cambiante por la aparición de nuevas tecnologías.
Si bien, existen indicadores que permiten analizar la salud de los proyectos de desarrollo estándar [1, 2], en el caso de los proyectos donde las herramientas de Inteligencia Artificial generativa [3] juegan un rol protagónico, es necesario establecer nuevas formas de medir [4] y controlar su evolución.
En este contexto, el presente trabajo tiene como objetivo aportar experiencias de campo considerando cómo se afectan los procesos de desarrollado cuando los mismo se realizan con el apoyo de Chats de Inteligencia Artificial Generativa. Lo cual servirá de base para la creación y/o ajuste de métricas.

Introducción

En la actualidad la IA ha cobrado un papel fundamental en la gestión y desarrollo de software [5]. Herramientas basadas en IA pueden asistir en la redacción de documentos, la generación de código, la creación de diagramas y la identificación de errores, optimizando cada etapa del proceso. Sin embargo, aunque sus beneficios son

significativos, aún requiere la intervención humana para validar y ajustar sus respuestas.

En este contexto se desarrolló una herramienta software que permite procesar una lista de acciones proporcionada por el usuario, descargar datos históricos de precios (mediante un proceso de Web Scrapy [6]), calcular los indicadores RSI (Relative Strength Index (Índice de Fuerza Relativa)) [7] y generar recomendaciones sobre posibles oportunidades de compra y/o venta.

El mismo fue elegido ya que combina diferentes tipos de programación y habilidades para su desarrollo. Como se observa en la Figura 1, intervienen varios componentes en el desarrollo de este sistema.

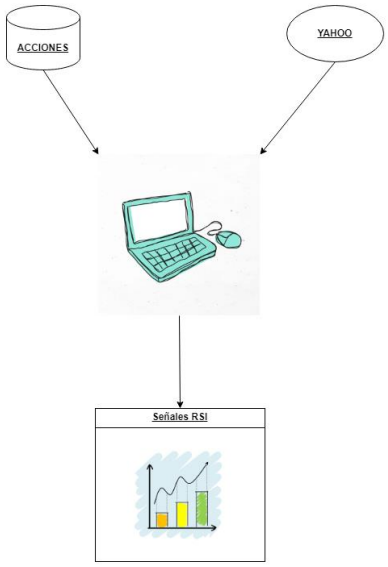


Figura 1: Esquema general del sistema de análisis bursátil

Líneas de Investigación y Desarrollo

El desarrollo de software asistido por IA generativa es una de las tendencias actuales de mercado [8,9], en la cual se espera mejorar no solo los tiempos sino también la calidad del producto a desarrollar.

Ahora bien, así como hay muchos trabajos de éxito relacionado con este tipo de herramientas, también existen expectativas de parte de los usuarios o sponsor de los proyectos que no se condicen con la realidad, por ello es fundamental establecer nuevas formas de medición y seguir los procesos de desarrollo que permitan generar expectativas viables respecto de las mejoras en los procesos de desarrollo.

El presente desarrollo servirá de base para establecer similitudes y diferencias en la forma de medir un desarrollo clásico y otro asistido por IA. Viendo, por ejemplo, como es la curva de aprendizaje de los analistas y desarrolladores, cual es el nivel de expertiz requerido de los analistas y desarrolladores en un entorno asistido, como se afectan los tiempos de proyecto y la calidad del mismo, entre otros aspectos. Se espera desarrollar nuevos trabajos de este tipo en 2025.

Resultados obtenidos

A continuación, se presentan los resultados obtenidos en cada una de las etapas del ciclo de vida desarrolladas:

Primera etapa: Especificación de Requisitos de Software

En esta etapa el ChatGPT resultó útil para organizar la información y generar un primer borrador de los requisitos. Sin embargo, las respuestas iniciales fueron limitadas en cantidad de requisitos, y al repetir consultas, las nuevas respuestas no siempre consideraban el contexto anterior.

Observaciones

> *ChatGPT resultó útil para organizar la información y generar un primer borrador de las especificaciones de requisitos. Sin embargo, las respuestas iniciales fueron limitadas en cantidad de requisitos, y al repetir consultas, las nuevas respuestas no siempre consideraban el contexto anterior.*

Segunda etapa: Análisis de requerimientos

Con los requisitos validados, se utilizó ChatGPT para generar el documento de casos de uso. La IA logró estructurar los casos de manera adecuada, aunque al solicitar diagramas, las imágenes generadas no fueron las correctas. La mejor solución fue utilizar herramientas como PlantUML o Mermaid para transformar el código proporcionado por ChatGPT en diagramas precisos.

Observaciones

La IA demostró ser eficiente en la redacción de casos de uso, pero al generar diagramas, es recomendable validar la sintaxis y adaptar el código según la herramienta utilizada.

Tercera etapa: Codificación del software

Para la generación de código, se probaron tres enfoques:

1. Proporcionar toda la documentación,
2. Proporcionar solo los requisitos clave (como los Requisitos Funcionales (RF) y Requisitos No Funcionales (RNF)),
3. Realizar consultas específicas sobre cierto bloque de código.

Los mejores resultados se obtuvieron con las consultas detalladas sobre funciones concretas, como la implementación de interfaces. Además, la IA fue muy útil para la resolución de errores y la instalación de librerías.

Observaciones

Si bien el enfoque más efectivo fue realizar consultas específicas, se ha utilizado el segundo enfoque para empezar el proyecto, ya que proporciona buena base y poco detalle (cabe resaltar que para este proyecto no se utilizó más de 15 RF). La IA facilita la generación de código, pero la intervención humana sigue siendo esencial para adaptar y optimizar las soluciones propuestas.

Cuarta etapa: Testing

Para la fase de pruebas, el ChatGPT generó los casos de prueba de manera estructurada y categorizada por casos de uso. Las respuestas fueron precisas y no requirieron modificaciones significativas. Sin embargo, la documentación de evidencias de prueba no pudo ser automatizada, ya que implicaba capturas de pantalla del software desarrollado.

Observaciones

La IA resultó efectiva para la generación de casos de prueba, pero su utilidad en la documentación de evidencias es limitada, dado que esta requiere información visual.

En base a la experiencia relevada se generará un conjunto de métricas e indicadores, los cuales serán validados en futuros trabajos de campo.

Formación de Recursos Humanos

El equipo de trabajo está formado por tres doctores, un magister (en proceso de doctorarse) y un conjunto de estudiantes adscriptos que rotan a lo largo del proyecto.

Se prevé desarrollar tesis de grado y trabajos profesionales vinculados al tema central del proyecto por estudiantes de la Universidad Nacional del Oeste. El equipo ya ha presentado algunos trabajos relacionados a la presente línea de investigación.

Referencias bibliográficas

[1] Metodologías de Ingeniería Informática; Ochoa, A. Fernández, E., Britos, P., García-Martínez, R. 2008. Editorial Nueva Librería. 869 páginas. ISBN 978-987-1104-54-3

[2] ComputerWeekly. (2020). "23 métricas de desarrollo de software que monitorear hoy". Recuperado de: https://www.computerweekly.com/es/consejo/23-metricas-de-desarrollo-de-software-que-monitorear

[3] https://www.shutterstock.com/es/blog/que-es-la-ia-generativa. Vigente al 18/006/2024

[4] María Alejandra Ochoa, Carlos Gustavo Lopez, Enrique Fernández. 2024. Impulsando La Productividad Y Calidad Del Software Con Inteligencia Artificial. 7° Congreso Argentino de Ingeniería / 18° Congreso Argentino de Enseñanza en Ingeniería (CADI / CAEDI)

[5] Coutinho, M., Marques, L., Santos, A., Dahia, M., França, C., & de Souza Santos, R. (2024). "The Role of Generative AI in Software Development Productivity: A Pilot Case Study". *arXiv preprint arXiv:2406.00560*. Disponible en: https://arxiv.org/abs/2406.00560

[6] Kouzis-Loukas, D. (2016). *Learning Scrapy*. Packt Publishing.

[7] Wilder, J. W. (1978). *New Concepts in Technical Trading Systems*. Trend Research

[8] Monteiro, M., Castelo Branco, B., Silvestre, S., Avelino, G., & Valente, M. T. (2023). "End-to-End Software Construction using ChatGPT: An Experience Report". *arXiv preprint arXiv:2310.14843*.

[9] Li, Y., Shi, J., & Zhang, Z. (2023). "A Novel Approach for Rapid Development Based on ChatGPT and Prompt Engineering". *arXiv preprint arXiv:2312.13115*.

Visión por Computadoras con Aprendizaje Profundo

Franco Ronchetti[1,2], Facundo Quiroga[1,2], Gastón Rios[1,3], Pedro Dal Bianco[1,3], Santiago Ponte Ahon[1], Oscar Stanchi[1], Juan Seery[1], Yael Aidelman[4], Roberto Gamen[4], Nicolas Francisco Abbate[5], Laura Lanzarini[1], Waldo Hasperué[1]

[1] Instituto de Investigación en Informática LIDI, Facultad de Informática, Universidad Nacional de La Plata, Argentina.*
[2] Comisión de Investigaciones Científicas de la Pcia. De Bs. As. (CICPBA)
[3] Becario postgrado UNLP
[4] Instituto de Astrofísica de La Plata, Universidad Nacional de La Plata, Argentina.
[5] Centro De Estudios Distributivos, Laborales Y Sociales, Universidad Nacional de La Plata, Argentina
* Centro asociado de la Comisión de Investigaciones Científicas de la Pcia. De Bs. As. (CIC)

Contacto: fronchetti@lidi.info.unlp.edu.ar

CONTEXTO

Esta presentación corresponde a algunas de las tareas de investigación que se llevan a cabo en el III-LIDI en el marco del proyecto "Inteligencia de Datos. Técnicas y Modelos de Machine Learning" perteneciente al Programa de Incentivos (2023-2026).

RESUMEN

Esta área de investigación se enfoca en el análisis y creación de sistemas y modelos basados en Deep Learning para abordar desafíos en el campo de la Visión por Computadora, integrando tanto métodos clásicos de Aprendizaje Automático, como enfoques avanzados de Aprendizaje Profundo mediante el uso de Redes Neuronales Convolucionales (CNNs) y arquitecturas basadas en Transformers. En este marco, los autores de este estudio implementan y adaptan dichas metodologías para resolver diversos problemas específicos:

- Interpretabilidad de modelos de Redes Neuronales, con énfasis en modelos post-hoc.
- Análisis y procesamiento de datos astronómicos, enfocados en la recuperación de trabajos de observación históricos.
- Estimación de ingreso per cápita utilizando imágenes satelitales.

Palabras clave: Redes Neuronales, Redes Convolucionales, Redes Recurrentes, Datos Astronómicos, Imágenes Satelitales, RISE, Interpretabilidad.

1. INTRODUCCION

El Instituto de Investigación en Informática LIDI (III-LIDI) de la UNLP tiene una larga trayectoria en el estudio, investigación y desarrollo de Sistemas Inteligentes basados en distintos métodos de Aprendizaje Automático y Redes Neuronales.

Como resultado de estas investigaciones se han diseñado e implementado técnicas originales aplicables para resolver diferentes problemas de Visión por Computadora utilizando técnicas de Aprendizaje Profundo. Particularmente, en la actualidad se están desarrollando los siguientes temas:

1.1. Interpretabilidad de modelos de Redes Neuronales.

Las redes neuronales son modelos tradicionalmente considerados como de caja negra. En los años recientes, se han realizado varios esfuerzos para comprender su funcionamiento de forma tal que el mismo sea más predecible o modulable. Uno de los algoritmos Post -hoc más utilizados es RISE. RISE (Randomized Input Sampling for Explanation) es un método de interpretabilidad para modelos de aprendizaje automático de caja negra, especialmente aplicado a datos de imágenes. Genera mapas de importancia (*heatmaps*) que destacan las regiones de una imagen más relevantes para la predicción del modelo. Para ello, RISE oculta aleatoriamente partes de la imagen con parches y evalúa cómo cambian las predicciones. Sin embargo, su enfoque original tiene limitaciones, como un alto costo

computacional y el uso de parches negros, que pueden distorsionar los resultados.

Para abordar estos desafíos, se proponen dos nuevas versiones del algoritmo: C-RISE y CB-RISE [2]. C-RISE incorpora un mecanismo de detección de convergencia basado en el algoritmo de Welford, que reduce la carga computacional al detener los cálculos una vez que el mapa de importancia se estabiliza. Por su parte, CB-RISE introduce el uso de máscaras difuminadas como perturbaciones, equivalentes a aplicar ruido gaussiano, en lugar de parches negros. Esto permite una representación más precisa del proceso de toma de decisiones del modelo. Los resultados experimentales demuestran la eficacia de estas mejoras, logrando una mayor calidad en los mapas de importancia generados y un *speedup* cercano a 3. Estas contribuciones refuerzan la utilidad de RISE como herramienta de interpretabilidad en aplicaciones de visión por computadora. Todos los experimentos fueron llevados a cabo con el *dataset* ImageNet con una arquitectura VGG16.

Por otro lado, se realizó una implementación de los diferentes modelos de RISE en la librería de interpretabilidad de modelos Captum de PyTorch [1]. Esto permite que cualquier investigador pueda evaluar de forma rápida y sencilla sus modelos.

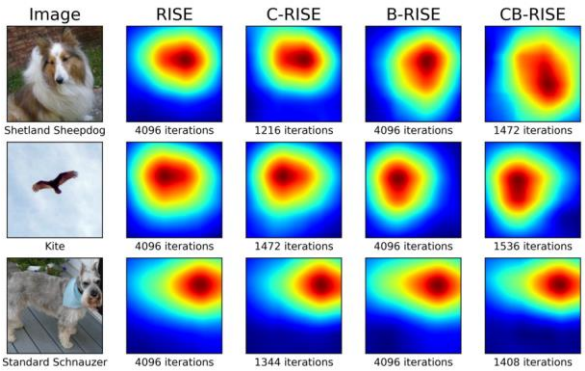


Figura 1. Variantes del algoritmo RISE utilizando el estándar de 4096 máscaras, de tamaño 4x4.

1.2 PlateUNLP. Sistema de digitalización y análisis de placas espectrográficas.

La Facultad de Ciencias Astronómicas y Geofísicas de la Universidad Nacional de La Plata posee 15,000 placas de vidrio con registros espectroscópicos que contienen datos astronómicos únicos y valiosos.

Desde 2019, el proyecto Recuperación del Trabajo Observacional Histórico (ReTrOH)[1] se ha dedicado a la digitalización estas placas espectroscópicas. El procesamiento manual de cada placa es un proceso complejo, propenso a errores y que requiere varias horas de trabajo. Para agilizar esta tarea y reducir los errores, se formó un equipo multidisciplinario de astrónomos e integrantes del III-LIDI que desarrolló PlateUNLP [3], un sistema que combina algoritmos de procesamiento de señales y modelos de visión por computadora. Este software asiste al usuario en el proceso de digitalización, permitiendo la corrección de predicciones automáticas y minimizando la necesidad de intervención manual. PlateUNLP detecta e individualiza automáticamente cada espectro registrado en las placas y agrega los metadatos correspondientes. El sistema está disponible públicamente bajo una licencia de código abierto[2]. Aunque fue diseñado para el proyecto ReTrOH, puede adaptarse para facilitar la digitalización de otras colecciones de placas astronómicas en el mundo

Un desafío particular de este proceso es la calibración de longitud de onda de las lámparas de comparación que poseen los espectros. Ya que no se dispone de los metadatos ni de las lámparas físicas originales, obliga a realizar un enfoque semi-automatizado basado en datos simulados. Sin embargo, estos datos simulados difieren significativamente de las observaciones reales debido a imperfecciones en las lámparas y el espectrógrafo, así como a picos teóricos que no siempre se observan en la práctica.

Como parte del proceso completo de extracción del espectro, se realizó un pipeline de calibración de longitud de onda que utiliza

[1] Proyecto ReTrOH: https://retroh.fcaglp.unlp.edu.ar/

[2] PlateUNLP: https://github.com/midusi/PlateUNLP

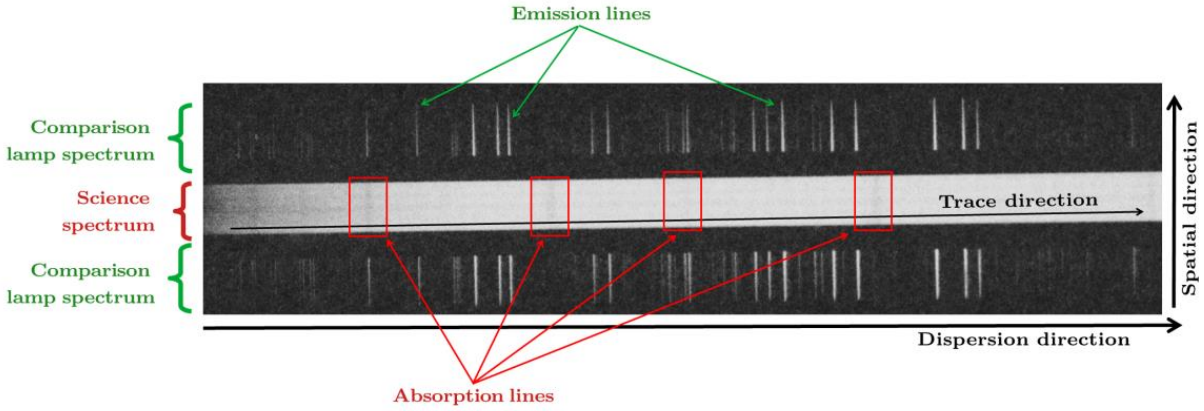


Figura 2. Una placa espectrográfica. La banda blanca central corresponde al espectro 2D científico. Los dos segmentos verticales brillantes ubicados en la parte superior e inferior de la imagen representan los dos espectros 2D de las lámparas de comparación.

Dynamic Time Warping (DTW) para comparar automáticamente las muestras con los datos simulados [4]. El mejor modelo logra un 93% de precisión promedio en la métrica Intersection-over-Union (IoU) sobre un conjunto de 32 placas calibradas manualmente, demostrando su efectividad para automatizar este proceso.

1.3 Estimación geográfica del ingreso per cápita del AMBA utilizando imágenes satelitales

El acceso a datos socioeconómicos desagregados y actualizados es crucial para la formulación y evaluación de políticas públicas efectivas. Sin embargo, la falta de disponibilidad y resolución de estos datos suele ser un obstáculo.

Una de las líneas de investigación en las que se está trabajando propone una metodología innovadora para estimar el ingreso per cápita a nivel altamente desagregado (50x50 metros) utilizando imágenes satelitales de alta resolución y una red neuronal convolucional (EfficientNetV2) [5]. A partir de datos censales de 2010 e imágenes satelitales de 2013 del Área Metropolitana de Buenos Aires (AMBA), el modelo logra predecir el ingreso con un alto nivel de precisión, alcanzando un R^2 del 85.9% en el conjunto de prueba. Este enfoque supera las limitaciones de resolución de los datos censales tradicionales, que solo ofrecen información a nivel de radio censal, y permite capturar patrones espaciales del ingreso con un detalle sin precedentes.

El trabajo desarrollado demuestra que el modelo es capaz de identificar áreas con diferentes niveles de ingreso, incluyendo asentamientos informales y barrios cerrados, lo que sugiere que las características visuales de las imágenes satelitales están fuertemente correlacionadas con el bienestar socioeconómico. Además, se generó un mapa de ingreso per cápita a nivel de grilla, que ofrece una visión detallada de la distribución espacial del ingreso en el AMBA. Este mapa no solo mejora la resolución de los datos censales, sino que también permite analizar dinámicas socioeconómicas a escalas más finas.

Finalmente, se evaluó la consistencia del modelo a nivel municipal, comparando indicadores clave como el ingreso medio, el coeficiente de Gini y la proporción de hogares bajo la línea de pobreza [6]. Aunque el modelo tiende a subestimar ligeramente los niveles absolutos de ingreso, captura correctamente las tendencias y el ordenamiento de los municipios según su nivel de ingresos. Este estudio destaca el potencial del enfoque para mejorar la formulación y evaluación de políticas públicas, ofreciendo herramientas más precisas y desagregadas para la toma de decisiones. Además, la metodología representa un avance en la capacidad

predictiva y la resolución espacial de indicadores socioeconómicos, sentando un precedente para futuras investigaciones en este campo.

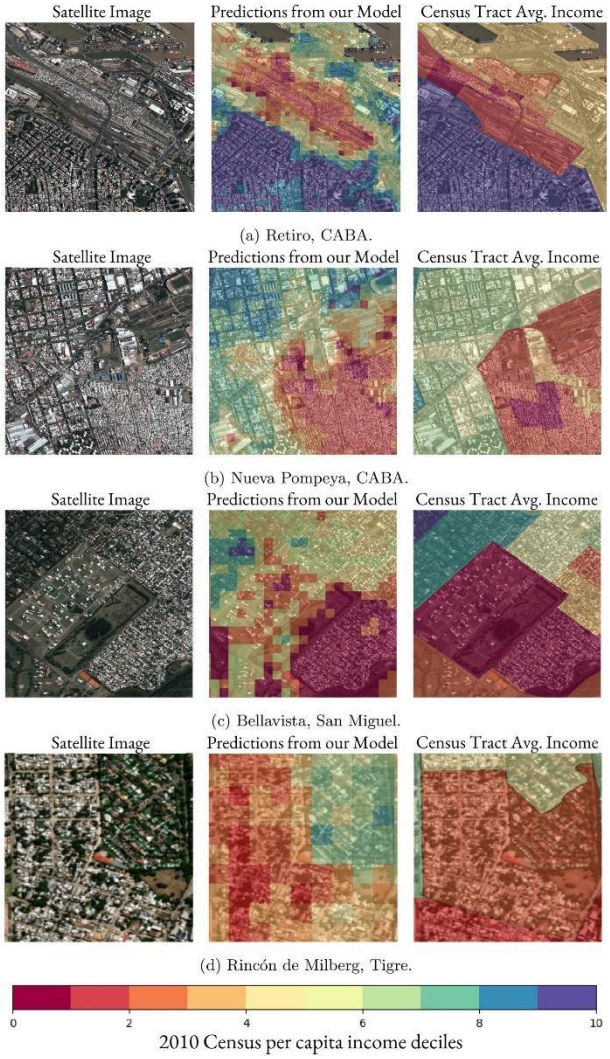


Figura 3. Casos de estudio particulares de predicción de ingreso per cápita.

2. LINEAS DE INVESTIGACIÓN Y DESARROLLO

- Redes neuronales profundas, convolucionales y transformers.
- Herramientas de interpretabilidad para visión por computadora.
- Variantes del algoritmo RISE para interpretabilidad de modelos de Redes Convolucionales.
- Calibración en longitud de onda de lámparas de comparación de espectros estelares.
- Predicción de ingreso per cápita usando imágenes satelitales.

3. RESULTADOS OBTENIDOS/ESPERADOS

- Nuevas variantes del algoritmo RISE para interpretabilidad de modelos.
- Modelo predictor de ingreso per cápita para el AMBA.
- Algoritmo de alineamiento de lámparas de comparación en imágenes de espectros históricos basado redes neuronales.
- Software de asistencia a la digitalización de placas espectrográficas antiguas.

4. FORMACIÓN DE RECURSOS HUMANOS

El grupo de trabajo de la línea de I/D aquí presentada está formado por: 3 profesores con dedicación exclusiva, un JTP dedicación exclusiva, 2 investigadores CIC-PBA, 2 becarios de posgrado de la UNLP con dedicación docente. Actualmente se están desarrollando 2 tesis de doctorado en este tema y 2 de maestría. También participan en el desarrollo de las tareas becarios y pasantes del III-LIDI.

5. REFERENCIAS

[1] Stanchi O., Ronchetti F., Quiroga F. The Implementation of the RISE Algorithm for the Captum Framework. Cloud Computing, Big Data & Emerging Topics. JCC-BD&ET 2023. Communications in Computer and Information Science, vol 1828. Springer, Cham.

[2] Stanchi, O., Ronchetti, F., Dal Bianco, P., Rios, G., Ponte Ahon, S., Hasperué, W., Quiroga, F. CB-RISE: Improving the RISE Interpretability Method Through Convergence Detection and Blurred Perturbations. Cloud Computing, Big Data and Emerging Topics. JCC-BD&ET 2024.

Communications in Computer and Information Science, vol 2189. Springer, Cham.

[3] Ponte Ahón, S.A., Aidelman, Y., Cidale, L., Gamen, R., Seery, J., Quiroga, F., Ronchetti, F., Hasperué, W., Iannuzzi, M., Peralta, R., López, M., Bariviera, A. PlateUNLP: An Open-Source Solution for Automating the Digitization of Historical Astronomical Plates. Eurographics Workshop on Graphics and Cultural Heritage 2024.

[4] Ponte Ahón, S.A., Seery, J., Quiroga, F., Ronchetti, Stanchi, O., Dal Bianco, P., Hasperué, W., Aidelman, Y., Gamen, R. Wavelength Calibration of Historical Spectrographic Plates with Dynamic Time Warping. Cloud Computing, Big Data and Emerging Topics. JCC-BD&ET 2024. Communications in Computer and Information Science, vol 2189. Springer, Cham.

[5] Abbate, N. F., Gasparini L., Ronchetti, F., Quiroga, F. High-Resolution Income Estimates Using Satellite Imagery: A Deep Learning Approach Applied in Buenos Aires. 2024 Latin American Computer Conference (CLEI), Buenos Aires, Argentina, 2024.

[6] Abbate, N.F. Mapeando el ingreso del Área Metropolitana de Buenos Aires en Alta Resolución: Un enfoque basado en Redes Neuronales aplicadas a Imágenes Satelitales. Tesis de Magister. 2024. Universidad Nacional de La Plata.

· ARSO ·

Arquitectura, Redes y Sistemas Operativos

Dockerización de Servidores SCADA: Ciberseguridad Industrial

Valentin Torassa Colombero; Santiago Enrique Roatta; Maria Eugenia Casco

CAETI – Universidad Abierta Interamericana
DV. Montes de Oca 725 – Buenos Aires - Argentina
{Valentin.TorassaColombero, Santiago.Roatta, MariaEugenia.Casco}@uai.edu.ar

RESUMEN

Los sistemas SCADA (Supervisory Control and Data Acquisition) se utilizan en la gestión y supervisión de procesos industriales. No obstante, su creciente conectividad a redes externas para gestión y administración remotas, los exponen a amenazas cibernéticas. En este contexto, la tecnología de contenedores con Docker se presenta como una solución capaz de mejorar tanto la seguridad como la eficiencia de los servidores SCADA, permitiendo un desarrollo, despliegue y mantenimiento más ágil de sus componentes.

El presente trabajo aborda la contenerización de sistemas SCADA con Docker, describiendo los principios básicos de esta tecnología y detallando cómo su aislamiento de procesos, escalabilidad y consumo eficiente de recursos contribuyen al desarrollo de infraestructuras industriales seguras y al acortamiento de los ciclos de actualización y mantenimiento**.** Presentando las ventajas y desafíos de esta aproximación, con el objetivo de reducir riesgos y asegurar la continuidad operativa.

Palabras clave: SCADA, Docker, Contenedores, Ciberseguridad Industrial, Infraestructura Crítica

CONTEXTO

Los proyectos radicados en el CAETI se clasifican en tres líneas prioritarias Automatización y Robótica, Ingeniería de Software y Sociedad del Conocimiento y Tecnologías Aplicadas a la Educación. Este proyecto se enmarca en la rama de Ingeniería de Software y más específicamente en la subdivisión de Ciberseguridad, conceptos y aplicaciones.

1. INTRODUCCIÓN

La creciente digitalización de los procesos industriales, impulsada por la denominada Industria 4.0, ha traído consigo avances en eficiencia, monitoreo en tiempo real y automatización [1]. De mano de estos cambios, los sistemas SCADA han ganado gran relevancia para la supervisión y el control de maquinaria, sensores y dispositivos diseñados para plantas de producción o redes de servicios.

La elevada interconexión expone estos sistemas a riesgos de seguridad, haciéndolos vulnerables a ciberataques como ransomware industrial (por ejemplo, LockerGoga) y a la explotación de protocolos no seguros (MODBUS, DNP3) [2]. De acuerdo con el NIST SP 800-82, el 67% de las brechas en entornos OT se debe a configuraciones deficientes y falta de aislamiento de componentes [3]. Asimismo, el acceso remoto no autorizado puede utilizarse para manipular procesos críticos, lo que se ha convertido en uno de los principales objetivos de los hackers, especialmente aquellos contratados por Estados para acciones ligadas a la ciberguerra. [4]

Asegurar los entornos SCADA para evitar interrupciones en la producción protege la integridad de activos de gran valor. Es en este

punto Docker —una plataforma de virtualización y solución basada en contenedores— ofrece varias ventajas aplicables a estos sistemas: aislamiento de aplicaciones, rapidez de despliegue y escalabilidad [5]. A diferencia de las máquinas virtuales tradicionales, los contenedores permiten empaquetar servicios con sus dependencias de forma ligera y segura siendo funcionalmente como una virtualización, pero optimizando recursos y reduciendo la superficie de ataque. [6]

El objetivo de esta investigación es presentar los beneficios de la contenerización de servidores SCADA utilizando Docker y, a su vez, ofreciendo una solucion para incrementar la protección de infraestructuras industriales críticas. Se revisarán los principios fundamentales de Docker, sus ventajas [5], y se explorarán las principales configuraciones de seguridad para mitigar vulnerabilidades, enfocado en entornos industriales donde la fiabilidad, contingencia y resiliencia son indispensables. [3]

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

La investigación se articula en tres líneas fundamentales, cada una enfocada en desarrollar los conceptos clave para el desarrollo y adaptación de la tecnología de contenedores para la industria. En la primera, se realiza un breve repaso del concepto de contenedores y la aplicación de Docker en entornos industriales. En la segunda línea, se profundiza en la contenerización de SCADA, incluyendo la arquitectura de estos sistemas y el despliegue con Docker. Finalmente, se analizan los beneficios y desafíos de esta integración.

2.1 Conceptos de Contenedores y Docker para la Industria 4.0:

Los contenedores son entornos de ejecución autocontenidos que encapsulan una aplicación junto con sus dependencias (librerías, configuraciones, etc.) en una sola unidad [5]. Esto difiere de las máquinas virtuales (VM), donde se virtualiza todo un sistema operativo completo incluidas todas sus capas. [6]

Al compartir el kernel del sistema operativo host, los contenedores resultan más livianos y consumen menos recursos que las VM [6]; Cada contenedor opera de forma independiente, lo que minimiza la interferencia e interdependencia entre servicios; Finalmente, gracias a la estandarización de las imágenes Docker, el mismo contenedor puede desplegarse en diferentes hosts (locales o en la nube) sin modificar su configuración [5].

2.2 Arquitectura y Contenerización SCADA

En la contenerización de SCADA, resulta fundamental conocer en detalle la arquitectura tradicional que respalda estos sistemas industriales. Usualmente, dicha arquitectura integra uno o varios servidores principales, dispositivos PLC (Programmable Logic Controllers) encargados de la ejecución de lógicas de control, y HMI (Human Machine Interface) que permiten la interacción de los operadores con el proceso. A esto se suman otros componentes como historiadores de datos (data historians) y pasarelas de comunicación que hacen uso de protocolos estándar (MODBUS, OPC, DNP3) o propietarios. [1]

Para llevar este ecosistema a un entorno contenerizado con Docker, se parte de la elaboración o adaptación de un Dockerfile que describa las dependencias del software SCADA (librerías, frameworks, servicios adicionales) y establezca parámetros de configuración —por

ejemplo, variables de entorno— esenciales para que el sistema funcione de manera consistente. [5]

En este proceso, resulta importante definir con precisión la exposición de puertos que los PLCs, u otros dispositivos externos necesitan para intercambiar datos con el servidor SCADA, así como la configuración de volúmenes que permitirán persistir información (por ejemplo, registros históricos de procesos o bitácoras de eventos). [1-5]

Es posible gestionar múltiples contenedores correlacionados —por ejemplo, el servidor SCADA, un sistema de bases de datos y una herramienta de monitoreo— a través de orquestadores como Docker Compose, logrando una implantación similar a la de un microservicio. [7]

Permitiendo replicar el entorno en distintas etapas —desarrollo, pruebas y producción— sin temor a discrepancias en las versiones de librerías o configuraciones específicas, sobretodo en operaciones que deben mantener la continuidad de servicio. [5]

La flexibilidad resultante de la adopción de Docker posibilita la actualización modular de cada parte del sistema, ante fallas o necesidades de parcheo, basta con regenerar y desplegar contenedores concretos en lugar de interrumpir todo el conjunto, optimizando el ciclo de vida del sistema. [5]

2.3 Beneficios y Desafíos de la implementación:

En la implementación de sistemas SCADA en docker, emergen tanto beneficios como desafíos, desde el punto de vista de la eficiencia, varias pruebas realizadas en laboratorios simulados han demostrado que la virtualización por contenedores permite una reducción del consumo de recursos singnificativa en comparación con entornos basados en máquinas virtuales, gracias al uso compartido del kernel y a la eliminación de capas innecesarias de software.[6] Este enfoque también ofrece la posibilidad de aislar procesos, con ello, se facilita la adopción de procesos de despliegue continuo (CI/CD) y se acelera la restauración del servicio en caso de fallos. [5]

- **Ahorro de costos en infraestructura:** Al requerir menos espacio en disco y memoria, es posible destinar recursos adicionales a otros servicios, reduciendo la inversión en servidores físicos o la facturación en la nube. [5]

- **Mayor disponibilidad y tolerancia a fallos:** En entornos de pruebas, las imágenes de docker han permitido reponer servicios SCADA afectados por incidentes en cuestión de segundo, minimizando así la inactividad debida la poca demanda de recursos para levantar estos subsistemas. [5]

- **Refuerzo de la seguridad:** El aislamiento inherente a los contenedores reduce la superficie de ataque y permite aplicar controles granulares (como reglas de red internas o restricciones de privilegios), aumentando las defensas frente ciberataques. [7]

Por otro lado, no pueden obviarse los desafíos que acompañan a esta estrategia. La complejidad inicial de la curva de aprendizaje para los equipos de ingeniería y TI puede ser alta, dado que se requiere un buen entendimiento del ecosistema Docker (Dockerfiles, orquestadores, redes internas, Docker Hub) y de cómo integrar adecuadamente los distintos componentes SCADA en contenedores. [8]

Asimismo, en el caso de plataformas SCADA privativas, se presentan restricciones que impiden o dificultan la generación de imágenes Docker

oficiales o soporte nativo, lo cual obliga a adaptar manualmente el software y a enfrentarse a problemas de compatibilidad. Por otro lado, algunos entornos industriales requieren certificaciones regulatorias y validaciones de cumplimiento que a veces no contemplan la virtualización por contenedores. [9]

3. RESULTADOS ESPERADOS

Se prevé que esta investigación proporcione, gracias a extensiones y avances, una alternativa eficiente para la implementación y administración de sistemas SCADA mediante Docker, reforzando la seguridad y resiliencia de la infraestructura industrial. En particular, se espera:

- Exponer las ventajas de emplear contenedores en entornos SCADA, mostrando reducciones de costos, tiempos de inactividad y riesgos de seguridad.

- Ofrecer recomendaciones concretas para la industria, incluyendo el cumplimiento de normativas internacionales (IEC 62443, ISA99, etc.) y pautas de configuración adaptadas a entornos operativos.

- Proyectar líneas de investigación futura, como la integración de Kubernetes y herramientas de orquestación de contenedores, con miras a escalar y optimizar aún más los servicios.

Asimismo, se busca que los hallazgos de la investigación sirvan de base formativa, de modo que profesionales y estudiantes interesados en la ciberseguridad y la gestión de sistemas críticos puedan adoptar, adaptar y contribuir a las estrategias propuestas en diferentes escenarios productivos.

4. FORMACION DE RECURSOS HUMANOS

La iniciativa contribuye a la formación y especialización de futuros profesionales en áreas como la administración de contenedores, la ciberseguridad industrial y el cumplimiento de normativas internacionales. Fortaleciendo el vínculo entre la academia y la industria, permitiendo que los hallazgos obtenidos tengan un impacto real en el desarrollo de entornos SCADA.

BIBLIOGRAFIA

[1] Macaulay, T., & Singer, B. L. (2011). Cybersecurity for industrial control systems: SCADA, DCS, PLC, HMI, and SIS. CRC Press.

[2] Al-Hawawreh, M. (2022). Developing an effective detection framework for targeted ransomware attacks in brownfield industrial (Doctoral dissertation, UNSW Sydney).

[3] Hahn, A., & Govindarasu, M. (2011, July). An evaluation of cybersecurity assessment tools on a SCADA environment. IEEE Power and Energy Society General Meeting (pp. 1-6).

[4] Nicholson, A., Webber, S., Dyer, S., Patel, T., & Janicke, H. (2012). SCADA security in the light of Cyber-Warfare. Computers & Security, 31(4).

[5] Poulton, N. (2023). Docker Deep Dive (2025 Edition). Nigel Poulton Media.

[6] Potdar, A. M., Narayan, D. G., Kengond, S., & Mulla, M. M. (2020). Performance evaluation of docker container and virtual machine. *Procedia C.S*.

[7] Combe, T., Martin, A., & Di Pietro, R. (2016). To docker or not to docker: A security perspective. *IEEE Cloud Computing*, *3*(5), 54-62.

[8] Miell, I., & Sayers, A. (2019). *Docker in practice*. Simon and Schuster.

[9] Dolezilek, D., Gammel, D., & Fernandes, W. (2020). Cybersecurity based on IEC 62351 and IEC 62443 for IEC 61850 systems.

Experimentación de Automatización de Red utilizando tecnologías de programación y equipamiento de red MikroTik

Santiago Pérez, Higinio Facchini, Alejandro Dantiacq, Bruno Roberti, Matilde Césari, Gerardo Venier, Alejandro Silnik
CeReCoN (Centro de Investigación y Desarrollo en Computación y Neuroingeniería)
Facultad Regional Mendoza, Universidad Tecnológica Nacional
Rodriguez 273, Mendoza, Argentina - *0261-5244576*
(santiagocp,higiniofac,alejandrod)@frm.utn.edu.ar

Resumen

Una amplia variedad de empresas, corporaciones, CSPs y especialistas han enfatizado la dificultad de gestionar redes modernas, que introducen innovaciones tecnológicas de alto impacto, como cloud computing, movilidad, nuevos perfiles de tráfico, NFV, IoT, Big Data, entre otras. La automatización de redes es una metodología en la que los dispositivos de red físicos y virtuales se configuran, aprovisionan, administran y prueban automáticamente mediante software. Además, recientemente ha habido un aumento en la cantidad de herramientas que asisten en la automatización de redes. Ambos hechos han marcado un cambio en la forma en que los administradores construyen y administran las redes. Se espera, que progresivamente, esta tendencia alcance, también, a las organizaciones de menor envergadura. El objetivo principal de este trabajo es explorar cómo las tecnologías de programación y automatización, específicamente Python y Ansible, pueden ser utilizadas para implementar conceptos de la gestión de redes. Se examinará la forma en que estas herramientas permiten definir y configurar la infraestructura de red como código, tratando la infraestructura como si fuera una aplicación de software, lo que facilita su gestión y mantenimiento. Se pretende llevar a cabo diferentes casos de estudio y pruebas en entornos de laboratorio, para demostrar la viabilidad y eficacia de la infraestructura como código en redes. Se analizarán también los beneficios asociados a la automatización de tareas repetitivas, la reducción de errores humanos y la capacidad de escalar y desplegar configuraciones de red de manera rápida y consistente.

El personal principal relacionado a esta línea de investigación es docente de grado y posgrado de la Universidad Tecnológica Nacional, Facultad Regional Mendoza, y tesistas de posgrado en la Maestria y Especialización en Redes de Datos de la misma institución.

Palabras clave: Automatización de Red, Herramientas de Automatización, Redes de Datos.

Introducción

La automatización de redes se describe como "el proceso de automatizar la configuración, gestión y operaciones de una red informática. Es un término amplio que incluye una serie de herramientas, tecnologías y metodologías utilizadas para automatizar los procesos de red". Un informe reciente del MIT Technology Review, redactado conjuntamente con Ericsson, proporciona una definición más técnica de este concepto como "la eliminación de tareas manuales repetibles y su sustitución por tareas programadas automatizadas con el uso de software". Ejemplos de automatización de redes incluyen la configuración del servidor, la programación del mantenimiento y la adición o eliminación de servicios. La automatización de la red que va más allá de un único servidor o

servicio e implica la configuración de varias funciones de red virtualizadas, a menudo, implica organizar cuidadosamente la gestión del flujo de trabajo en toda la red. Un ejemplo concreto de esta metodología es la prueba exitosa de Vodafone de automatización total de sus servicios de conectividad de transporte.

Estado del Arte

Uno de los principales obstáculos para los administradores de redes en las organizaciones implica los crecientes costos de TI asociados con las operaciones de red. La creciente cantidad de datos, dispositivos y operaciones necesarias en dichos dispositivos ha comenzado a exceder las capacidades de TI. En algunos casos, los enfoques manuales resultan insuficientes o prácticamente imposibles. Como señala Cisco Systems, hasta el 95% de todos los cambios de configuración de la red se realizan manualmente, lo que genera costos operativos que pueden ser de 2 a 3 veces el costo de los activos de la red. Cada empresa debe evaluar su propia situación en un intento de mantenerse al día con los cambios que trae el mundo digital, y decidir si es hora de aumentar la automatización y gestionarla de forma remota y centralizada.

Las soluciones de automatización de redes pueden abordar una amplia gama de actividades directa o indirectamente, algunas de ellas son:

- Proporcionar dispositivos y servicios físicos o virtuales.
- Verificar los dispositivos y su configuración.
- Planificar escenarios y gestión del inventario de TI.
- Cumplir con las políticas y pautas de configuración de todos los dispositivos y servicios de red.
- Recopilar datos de red relacionados con dispositivos, sistemas de información, programas de sistemas, topología de red, tráfico y servicios en tiempo real.
- Analizar datos, incluidos análisis predictivos de IA y aprendizaje automático (ML).
- Actualizar con auditoría programas, incluida la reversión del software si es necesario.
- Reparar problemas de red de circuito cerrado.
- Poner a disposición información sobre incidencias, paneles, alertas y alarmas.
- Hacer cumplir las políticas de seguridad en mayor o menor medida.
- Supervisar la red y sus servicios, asegurando el cumplimiento de los acuerdos de nivel de servicio (SLA) entre proveedores y clientes.

Cabe señalar que la automatización tiene algunas características propias para estas actividades: 1) las herramientas de automatización pueden funcionar las 24 horas del día, los 7 días de la semana, sin interrupciones, lo que resulta en una mayor eficiencia; 2) La automatización ayuda a recopilar grandes cantidades de datos que, a su vez, pueden analizarse rápidamente para proporcionar información que pueda guiar un evento o proceso determinado; y 3) En diversas circunstancias, una herramienta de automatización de red inteligente puede alterar su comportamiento para lograr algún objetivo.

Las empresas que han sido pioneras en la adopción de la automatización han obtenido múltiples beneficios importantes:

- Costos más bajos: Debido a que la infraestructura subyacente es menos compleja, se necesitan menos horas de mano de obra para configurar, aprovisionar y administrar la red y sus servicios.
- Menor probabilidad de errores humanos: Se reducen los posibles errores asociados al funcionamiento manual, como errores de configuración o tipografía, entre otros. En consecuencia, se necesitan menos miembros del personal para resolver estos problemas.
- Mayor fuerza laboral estratégica: Al automatizar tareas repetitivas, las empresas aumentan su productividad, mejoran sus negocios e innovación y crean nuevas oportunidades de empleo para su fuerza laboral actual.

- Reducción del tiempo de inactividad de la red: Las empresas pueden alcanzar mayores niveles de disponibilidad de la red, con el consiguiente aumento de la productividad, mejorar su imagen corporativa y ofrecer servicios de mayor calidad.
- Desarrollo acelerado: Es posible adoptar nuevas aplicaciones nativas de la nube y procesos de flujo de trabajo DevOps a través de la gestión y seguridad de la red, que pueden integrarse fácilmente en los procesos de desarrollo de estas nuevas aplicaciones.
- Implementación acelerada: El tiempo de inicio de nuevas aplicaciones y servicios se puede reducir automatizando el aprovisionamiento y la gestión y seguridad de la red durante todo el ciclo de vida de las aplicaciones, en centros de datos y entornos de nube.
- Mayor conocimiento y control de la red: Lo que ocurre dentro de la red puede entenderse y analizarse mejor de forma global, promoviendo una evaluación más integral y la capacidad de autoadaptación cuando sea necesario.

Estado de Avance

El sector académico también juega un papel activo al realizar contribuciones significativas en redes definidas por software (SDN), IA, ML y simulación aplicadas a la automatización de redes. Esta sección presenta una selección de estudios que muestran los múltiples alcances y perspectivas que se encuentran en la investigación de la automatización de redes. Uno de estos estudios [1] presenta un modelo abstracto común para toda una red de transporte, la estandarización de las reglas de diseño e implementación y la configuración basada en intenciones. El documento antes mencionado afirma que este modelo debería capturar de forma autónoma cualquier cambio de estado entre las redes actuales y planificadas, y aceptar una variedad de estados de red futuros e implementaciones progresivas sin afectar la topología o red actual. Sin embargo, otro estudio [2] analiza el impacto del Internet de las cosas (IoT) y las tecnologías 5G en las redes actuales. Se considera que estas tecnologías han conducido a la expansión de la escala de la red y a nuevas aplicaciones. Se propone un Smart Integration Identifier Networking (SINET-I), destinado a mejorar la capacidad de automatización de la red y promover la colaboración en redes heterogéneas.

En [3] se presenta un enfoque alternativo, que surge de los lenguajes de programación utilizados en entornos y plataformas de automatización de redes. El estudio consistió en analizar, en un entorno experimental, cuál es el mejor método para mejorar la eficiencia de los scripts aplicados a los dispositivos de red. Además, se hizo una comparación entre el tiempo necesario para la configuración manual y la automática, y las diferencias resultantes en el rendimiento. Los autores diseñaron una topología de red que consta de 36 dispositivos Cisco, con diferentes versiones del sistema operativo de red (IOS) y concluyeron que el método automatizado reduce el tiempo requerido en más del 90%. En [4] se presenta un ángulo diferente, en el que SDN, telemetría de red en banda (INT) y análisis de datos se combinan para realizar un sistema de automatización de red de circuito cerrado. El diseño del plano de datos y del plano de control se demostró experimentalmente con un prototipo de sistema de red, que consta de seis paneles de distribución de energía solar. El estudio concluyó que la automatización de redes de circuito cerrado se puede lograr de manera efectiva.

Finalmente, los sistemas 5G, la automatización y la inteligencia en infraestructuras se analizan en [5], donde también se evalúa la necesidad de una mayor automatización en el mundo de las telecomunicaciones. En ese artículo, los autores evaluaron el alcance de la IA, el aprendizaje por refuerzo (RL) y el aprendizaje federado (FL) en las redes automatizadas del futuro.

El presente proyecto implica el afianzamiento del estudio de la Automatización de Red, iniciada en diversos proyectos internos de CeReCoN. Los integrantes del Proyecto son miembros del Grupo GRID ATyS, y han trabajado en distintos proyectos en el área de las TICs, en las temáticas de análisis de tráfico, redes avanzadas, Internet Version 6, modelación de redes Wi-Fi, ciberseguridad, IoT, SDN, etc. (en la Universidad Tecnológica Nacional).

Se aporta el conocimiento y experiencia de algunos de sus miembros en el tema de Automatización de Red, debido a sus condiciones de Doctor y Magister, y Especialistas. Además, varios miembros son docentes de grado y posgrado (en la Especialización y Maestría en Redes de Datos, modalidad a distancia), y por ser instructores de los cursos oficiales avanzados de Redes de Cisco de la carrera de CCNP y de CCNA (específicos de las temáticas de redes en todos los niveles del modelo OSI).

Finalmente, el Dr. Ing. Santiago Perez ha participado como expositor de la temática de Automatización Industrial en actividades extensionistas, y publicado resultados experimentales en informes técnicos internos.

Objetivos

Objetivo General:

El presente trabajo tiene como objetivo determinar experimentalmente el alcance y potencialidad de la automatización de red realizada sobre diferentes tipos de equipamiento MikroTik utilizando Python y Ansible.

Objetivos Específicos:

Para el presente proyecto se buscan los siguientes objetivos:

- Determinar los escenarios de prueba o experimentación para el análisis comparativo.
- Establecer la modalidad de comparación con la configuración manual versus la automatización de red.
- Determinar las configuraciones necesarias utilizando Ansible.
- Determinar las configuraciones necesarias utilizando Python.

Metodología

Para el desarrollo del proyecto se ha previsto las siguientes tareas:

Tarea 1: Recopilar de información y estudio de los alcances, ventajas y desventajas de la automatización de red.

Tarea 2: Compilar trabajos de investigación sobre la temática, usando Python y/o Ansible determinando escenarios experimentales, herramientas de simulación, tráficos utilizados, métricas medidas, resultados y conclusiones.

Tarea 3: Establecer mecanismos comparativos de los trabajos de investigación compilados, usando cuadros, índices, ponderaciones, etc. respecto a las configuraciones manuales.

Tarea 4: Definir un escenario de experimentación general, y particulares para Ansible sobre equipamiento MikroTik.

Tarea 5: Definir un escenario de experimentación general, y particulares para Python sobre equipamiento MikroTik.

Tarea 6: Construir tablas comparativas de métricas y alcances de la automatización entre Python y Ansible. Se construirán tablas y graficas comparativas de las prestaciones, alcances, o métricas para cada caso, según las combinaciones que permitan los diversos escenarios.

Tarea 7: Documentar.

Tarea 8: Redactar Informe Final.

Formación de Recursos Humanos

Mediante el presente proyecto se busca la formación de Recursos Humanos con el fin de:

- Aglutinar RRHH calificados en torno a estas nuevas tecnologías y crear las sinergias para facilitar su participación en proyectos e iniciativas nacionales.

- Facilitar la organización sostenida de seminarios científico-tecnológicos para la actualización de nuestro entorno y una visión prospectiva sobre la proyección de futuro de estas tecnologías.

Al efecto, se pone en conocimiento la propia publicación asociada a la temática [6].

Contribuciones Principales del Proyecto

Las contribuciones principales serán:

1. Por el uso de Ansible y Python como herramientas de automatización sobre Mikrotik: A mayor conocimiento de los integrantes del proyecto este es el primer trabajo que utilizará herramientas de automatización sobre equipamiento real, tipo MikroTik. Por lo tanto, las conclusiones serán un importante aporte para transferir la experiencia práctica que se obtenga.
2. Por el uso de nuevos escenarios más realistas de evaluación: Se pretende proponer escenarios sobre los que se analizará el comportamiento del equipamiento MikroTik, usando Python y Ansible.
3. Por la transferencia en posgrado: El Director y Codirector propuestos, y otros integrantes son actualmente docentes de cursos de posgrado en dichas carreras, que se dicta en la Facultad Regional Mendoza, lo que permite transferir todo el conocimiento científico y/o técnico a los estudiantes de dichas carreras.
4. Por la transferencia en revistas, congresos y workshops de divulgación: Además, debido a que se trata de un tema actual y de relevancia real en las redes de datos existentes en cualquier Institución, ente de Gobierno y/o empresarial, su transferencia en dichos ambientes, podrá ser inmediata, cumpliéndose las etapas de transferencia científica, tecnológica y productiva con la divulgación en medios adecuados, a saber:
   - Publicaciones en revistas de divulgación nacional e internacional.
   - Presentación en Congresos y Seminarios.
   - Vinculación con centros de estudio que trabajen el mismo tema, para cruzar conclusiones y verificar críticas.

Referencias

[1] T. Grahek, and A. Leong, "Network Automation for Design, Build and Operation at Scale," in 2019 Optical Fiber Communications Conference and Exhibition (OFC), San Diego, CA, USA, 2019.

[2] D. Chen, D. Gao, W. Quan, Q. Wang, G. Liu, and H. Zhang, "Promoting Network Automation for Heterogeneous Networks Collaboration," in 2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall), Victoria, BC, Canada, 2020.

[3] A. Mahmood Mazin, R. Ab Rahman, M. Kassim, and A. Razak Mahmud, "Performance Analysis on Network Automation Interaction with Network Devices Using Python," in 2021 IEEE 11th IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE), Penang, Malaysia, 2021.

[4] S. Tang, H. Liang, M. Wang, T. Li, and Z. Zhu, "Closed-loop Network Automation with Generic Programmable Data Plane (G-PDP)," in 2021 International Conference on Computer Communications and Networks (ICCCN), Athens, Greece, 2021.

[5] L. Militano, A. Zafeiropoulos, E. Fotopoulou, R. Bruschi, and C. Lombardo, "AI-powered Infrastructures for Intelligence and Automation in Beyond-5G Systems," in 2021 IEEE Globecom Workshops (GC Wkshps), Madrid, Spain, 2021.

[6] Pérez, Santiago; Facchini, Higinio; Dantiacq, Alejandro, Roberti, Bruno; Experimental Perfromance Contrast Between SDN and Tradicional Networks, IEEE CHILECON, Santiago, Chile, 06 al 09 de diciembre de 2021.

Investigación y Desarrollo sobre la enseñanza práctica de tecnologías de Internet de las Cosas

Gustavo Mercado, Raúl Moralejo, Ana Laura Diedrichs,
Mario Sebastián Tobar y Rodrigo Gonzalez

Universidad Tecnológica Nacional, Facultad Regional Mendoza, GridTICs,
Mendoza, Argentina.
{gmercado,mstobar,ana.diedrichs,rodralez}@frm.utn.edu.ar
romoralejo@gmail.com

RESUMEN

El presente proyecto de Investigación y Desarrollo se orienta a transformar la enseñanza práctica de tecnologías IoT mediante el diseño e implementación de un Laboratorio Remoto de Internet de las Cosas. Su objetivo es establecer las bases académicas y tecnológicas para la formación integral de estudiantes y docentes en sistemas "end to end", facilitando el acceso remoto a recursos de hardware, software y herramientas de análisis de datos. La metodología propuesta contempla la conversión de prácticas presenciales a un formato virtual, permitiendo la realización de experimentos en tiempo real y el monitoreo interactivo de dispositivos conectados. Asimismo, se desarrollarán métricas específicas para evaluar el impacto del aprendizaje no presencial, integrando evaluaciones continuas y retroalimentación entre los participantes. Con un enfoque en la escalabilidad, modularidad y sostenibilidad, el proyecto busca garantizar la actualización constante frente a la rápida obsolescencia tecnológica. Además, se analiza la posibilidad de extender este sistema a un ámbito industrial, ofreciendo soluciones a empresas y startups que requieran validar aplicaciones IoT sin incurrir en altos costos. En definitiva, esta iniciativa innovadora promueve la optimización de la enseñanza en ingeniería, fortaleciendo competencias técnicas y favoreciendo la colaboración interinstitucional para la formación de recursos humanos altamente calificados con alto impacto.

Palabras Claves: IoT, Laboratorio Remoto, Educación Virtual, Sistemas Embebidos

CONTEXTO

El presente artículo se vincula con el proyecto de investigación y desarrollo PID MCCEC375, «Investigación y desarrollo sobre la enseñanza práctica de las tecnologías de Internet de las Cosas - Remote Iot Lab», desarrollado por el grupo de investigación gridTICs de la Universidad Tecnológica Nacional, Facultad Regional Mendoza, Argentina.

1. INTRODUCCIÓN

El avance acelerado de las tecnologías de Internet de las Cosas (IoT) ha generado una transformación en diversos sectores, desde la industria y la agricultura hasta la educación. En este contexto, la formación de profesionales en ingeniería que dominen tanto los aspectos teóricos como los prácticos de los sistemas "end to end" de IoT se vuelve imprescindible. Este proyecto se orienta a establecer las bases académicas y tecnológicas para mejorar la enseñanza práctica en ingeniería, mediante la implementación de un Laboratorio Remoto de IoT que permita el acceso, monitoreo y análisis en tiempo real de dispositivos y sensores conectados a una red. La propuesta surge como respuesta a las limitaciones de los laboratorios tradicionales, en los cuales el acceso a recursos de hardware y software resulta costoso, restringido y sujeto a la rápida obsolescencia de los equipos.

La necesidad de contar con ambientes de aprendizaje que integren las ventajas de la virtualidad y la educación a distancia es cada vez más evidente, especialmente en escenarios donde las restricciones logísticas y económicas dificultan la adquisición y mantenimiento de equipos de laboratorio. La metodología convencional basada en la presencialidad se

enfrenta a retos importantes, como la disponibilidad limitada de dispositivos, los costos elevados de implementación y la dificultad para replicar de manera uniforme las condiciones de experimentación entre diferentes instituciones. Por ello, este proyecto propone la transformación de prácticas presenciales a un formato virtual remoto, garantizando que tanto estudiantes como docentes dispongan de una herramienta flexible y escalable para la realización de experimentos y validaciones.

La introducción de laboratorios remotos en el ámbito académico permite no solo optimizar los recursos disponibles, sino también fomentar la colaboración interinstitucional. El diseño modular del laboratorio propuesto posibilita la integración de diversas tecnologías de IoT, desde sensores y motes hasta gateways y sistemas de cloud computing, en una plataforma única que facilita el acceso a experimentos de gran complejidad. Esta característica es fundamental, ya que permite adaptar el sistema a las necesidades específicas de diferentes carreras y programas académicos, potenciando la formación integral de los futuros ingenieros. Además, el uso de herramientas de código abierto tanto en el software como en el hardware se alinea con la filosofía de transparencia y replicabilidad que caracteriza a las iniciativas innovadoras en el ámbito educativo.

El proyecto se fundamenta en un riguroso análisis del estado del arte, considerando tanto la evolución de los laboratorios remotos como las metodologías de enseñanza que han demostrado eficacia en ambientes virtuales. Estudios recientes han evidenciado que la combinación de recursos tecnológicos y estrategias pedagógicas innovadoras no solo mejora la calidad del aprendizaje, sino que también incrementa la motivación y autonomía de los estudiantes. En este sentido, el desarrollo de métricas específicas para evaluar el impacto del aprendizaje no presencial resulta un componente crucial, ya que permite comparar de manera objetiva el desempeño entre métodos tradicionales y nuevos enfoques didácticos. La evaluación continua, a través de herramientas de análisis de datos y encuestas periódicas, asegura la identificación temprana de oportunidades de mejora y la adaptación del sistema a las necesidades emergentes del entorno educativo.

El proyecto también aborda la problemática de la obsolescencia tecnológica, proponiendo un enfoque flexible y actualizable que permita la incorporación de nuevas herramientas y tecnologías a medida que estas evolucionen. La modularidad del laboratorio remoto es esencial para garantizar que, ante el rápido avance en el campo de IoT, se puedan realizar actualizaciones de hardware y software sin necesidad de una reestructuración completa del sistema. Esta característica es especialmente relevante en un entorno donde la innovación tecnológica es constante y las demandas del mercado laboral exigen profesionales capacitados en el uso de tecnologías de vanguardia.

Desde el punto de vista pedagógico, la implementación de este laboratorio remoto se concibe como una herramienta transformadora para la enseñanza de sistemas embebidos y de IoT. Al facilitar el acceso a un entorno de experimentación realista, el proyecto permite a los estudiantes adquirir competencias prácticas fundamentales para el desarrollo de soluciones tecnológicas complejas. La capacidad de programar, compilar y actualizar dispositivos en un entorno virtual no solo refuerza los conocimientos teóricos adquiridos en el aula, sino que también impulsa el desarrollo de habilidades críticas como el análisis, la resolución de problemas y la innovación. En este sentido, el laboratorio remoto se posiciona como un puente entre el conocimiento académico y la aplicación práctica en la industria, contribuyendo significativamente a la formación de recursos humanos altamente calificados.

Otro aspecto relevante de la propuesta es la posibilidad de ampliar el alcance del laboratorio a un ámbito industrial. El análisis preliminar del proyecto contempla la

transformación del laboratorio universitario en una plataforma de servicios para empresas y startups, lo que permitiría validar aplicaciones de IoT sin incurrir en los elevados costos asociados a la adquisición de equipos especializados. Esta extensión del proyecto no solo diversificaría las fuentes de financiamiento, sino que también promovería la transferencia de conocimiento entre el mundo académico y el empresarial, generando sinergias que favorezcan la innovación y el desarrollo tecnológico.

La implementación del laboratorio remoto se enmarca en un proceso meticuloso que abarca desde la definición de las especificaciones funcionales y técnicas hasta la realización de pruebas integrales de desempeño. El proyecto contempla una serie de actividades orientadas a transformar las prácticas presenciales en experiencias virtuales sin perder el rigor académico ni la calidad de la enseñanza. En este proceso, la colaboración entre docentes, investigadores y estudiantes es fundamental para asegurar que el laboratorio cumpla con los objetivos propuestos y se adapte a las necesidades del entorno educativo contemporáneo.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Para el desarrollo del presente proyecto se identifican cuatro líneas fundamentales:

Investigación del Estado del Arte y Tecnologías Emergentes
En esta primera etapa se recabará información exhaustiva sobre las soluciones existentes en laboratorios remotos de IoT, explorando las metodologías utilizadas en la enseñanza práctica de sistemas “end to end”. Se estudiarán plataformas de hardware y software, herramientas de virtualización, y estrategias de integración de sistemas que permitan el acceso remoto a dispositivos y la administración de datos en tiempo real. Este análisis servirá para identificar tendencias, limitaciones y oportunidades de mejora en el ámbito educativo y tecnológico.

Diseño de la Metodología y Especificaciones Técnicas
La segunda fase se centrará en el desarrollo de una metodología adaptada a la conversión de prácticas presenciales a un formato virtual remoto. Se definirán las especificaciones funcionales y técnicas que regirán el laboratorio, enfatizando la modularidad, escalabilidad y el uso de herramientas de código abierto. Asimismo, se establecerán los parámetros de interacción entre usuarios, dispositivos IoT y plataformas de análisis, garantizando la interoperabilidad y actualización constante frente a la rápida obsolescencia tecnológica.

Desarrollo e Implementación del Prototipo Funcional
En la tercera etapa se procederá a la integración de los componentes de hardware y software necesarios para crear un prototipo funcional del laboratorio remoto. Se desarrollarán interfaces de acceso, sistemas de monitoreo en tiempo real y mecanismos para la administración y control de recursos. Este prototipo servirá como base para validar la viabilidad técnica y pedagógica de la propuesta, permitiendo la realización de pruebas experimentales en entornos educativos reales.

Evaluación del Desempeño y Proyección a Entornos Industriales
Finalmente, se diseñarán e implementarán métricas de evaluación para analizar el impacto del laboratorio en el aprendizaje y su eficiencia operativa. Se realizarán pruebas de usabilidad, encuestas y análisis comparativos entre metodologías tradicionales y remotas. Esta línea también contempla el estudio de la viabilidad para adaptar la solución a un contexto industrial, ampliando su aplicabilidad a empresas y startups que requieran sistemas de IoT.

3. RESULTADOS ESPERADOS

Al término del proyecto se espera alcanzar los siguientes resultados:

1. Implementación exitosa de un Laboratorio Remoto de IoT que integre hardware y software, permitiendo la experimentación en

tiempo real.
2. Validación de una metodología innovadora para la transformación de prácticas presenciales a un formato virtual, garantizando calidad educativa.
3. Desarrollo de métricas de evaluación que demuestren el impacto del aprendizaje no presencial y faciliten la medición del desempeño.
4. Estudio de viabilidad para adaptar la solución a entornos industriales, posibilitando su replicación en empresas y startups.
Los resultados serán difundidos en conferencias y publicaciones especializadas.

Finalmente, los resultados obtenidos serán publicados tanto en conferencias internacionales de alta calidad con referato, como en revistas especializadas de alto impacto, parte del catálogo de Scopus y Web of Science.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo estará integrado por cinco investigadores destacados:

Ing. Gustavo Mercado (Director)
Profesor e investigador del grupo de Tecnologías de la Información y Comunicaciones, con amplia experiencia en la dirección de proyectos de investigación aplicada a IoT. Su trayectoria abarca el diseño e implementación de soluciones tecnológicas y la transformación de prácticas presenciales a entornos virtuales. Su rol es fundamental para coordinar la integración de hardware y software en el laboratorio remoto.

Dr. Raúl Omar Moralejo (Co-director)
Profesor e investigador con sólida experiencia en ingeniería y sistemas de información. Su expertise en la implementación de tecnologías y metodologías de enseñanza innovadoras le permite contribuir significativamente al desarrollo técnico y metodológico del proyecto.

Ing. Mario Sebastián Tobar (Investigador de Apoyo)
Especialista en redes y laboratorios remotos, el Ing. Tobar aporta su experiencia en el manejo de sistemas de IoT y la implementación de soluciones de monitoreo en tiempo real, asegurando la integración efectiva de los diversos componentes del laboratorio.

Ing. Ana Laura Diedrichs (Investigadora Formada)
Con formación en ciencia de datos y amplia experiencia en la aplicación de técnicas de análisis y modelado, la Ing. Diedrichs se enfoca en el desarrollo de métricas de evaluación y en la optimización de procesos de aprendizaje no presencial. Su participación fortalece la calidad y la relevancia de los resultados académicos.

Dr. Rodrigo Gonzalez (Investigador Formado)
Integrante activo del grupo de investigación, el Ing. González posee experiencia en el diseño y análisis de sistemas embebidos, aportando conocimientos en la implementación y verificación de soluciones IoT que complementan la visión integral del proyecto.

En el campo de la formación de recursos humanos se espera:
1. Iniciar en investigación a profesionales interesados en postularse a becas doctorales en el marco del proyecto.
2. Incorporar a alumnos y graduados de la UTN como becarios para fomentar la investigación aplicada.
3. Estimular a estudiantes de últimos años de carrera para desarrollar proyectos finales basados en esta innovadora propuesta.

5. BIBLIOGRAFÍA

(Giordano, 2018) Roberto Giordano Lerena, Sandra Cirimelo, Editores. Propuesta de estándares de segunda generación para la acreditación de carreras de ingeniería en la república argentina. Libro Rojo del CONFEDI, CONFEDI, Universidad FASTA Ediciones, 2018.

(Khaitan, 2014) Khaitan et al. "Design Techniques and Applications of Cyber Physical Systems: A Survey", IEEE Systems Journal, 2014.

(Orduña et al., 2020) Orduña, Pablo; Rodriguez-Gil, Luis; Angulo, Ignacio; Hernandez, Unai; Villar, Aitor; Garcia-Zubia, Javier. WebLabLib: New Approach for Creating Remote Laboratories, 2020.

(Zutin, 2018) Zutin, Danilo Garbi. «Online Laboratory Architectures and Technical Considerations». En Cyber-Physical Laboratories in Engineering and Science Education, editado por Michael E. Auer, Abul K.M. Azad, Arthur Edwards y Ton de Jong, pp. 5-16. Cham: Springer International Publishing, 2018.

(Tobar et al, 2020) M. S. Tobar, C. Taffernaberry, C. Perez and G. Mercado, "Laboratorio Remoto para Sistemas Embebidos", 2020 IEEE Congreso Bienal de Argentina (ARGENCON), Resistencia, Argentina, 2020, pp. 1-1, doi: 10.1109/ARGENCON49523.2020.9505398.

(Feisel & Rosa, 2005) Feisel, Lyle D., y Albert J. Rosa. «The Role of the Laboratory in Undergraduate Engineering Education». Journal of Engineering Education, 94(1) (enero de 2005): 121-130.

(Perez-Monte, C.F. et al. 2024). C F. Perez-Monte, G Mercado, C Taffernaberry, A L Diedrichs, F Piccoli, M S Tobar, M Ledda, R Moralejo y R Gonzalez. “Embedded Systems Programming Through eUCCvm, an Integrated Educational Tool.” In: Pesado, P., Panessi, W., Fernández, J.M. (eds) Computer Science – CACIC 2023. CACIC 2023. Communications in Computer and Information Science, vol 2123. Springer, Cham. https://doi.org/10.1007/978-3-031-62245-8_5

(Barrionuevo et al., 2018) Barrionuevo, M., Gil, C., Giribaldi, M., Suarez, C., Taffernaberry, C. Virtualization in Education: Portable Network Laboratory. In: De Giusti, A. (eds) Computer Science – CACIC 2017. Communications in Computer and Information Science, vol. 790. Springer, Cham. Online ISBN 978-3-319-75214-3, 2018.

(Conectar, 2011) Fontdevila, Pablo A. "Estudio de caso: Conectar igualdad." Revista Iberoamericana de Ciencia, Tecnología y Sociedad-CTS, 6.18, 2011.

(Zimmaro, s.f.) Zimmaro. https://intranetsp.bournemouth.ac.uk/documentsrep/Writing%20Good%%, s.f.

(Ortiz et al., 2004) Ortíz Maldonado, Alberto; Ortega, Alejandro; Day, Mario; Cicero, Aldo Roberto; Caretta, Adriana I. «Experiencia de evaluación integradora grupal en Meteorología Agrícola». Revista de la Facultad de Ciencias Agrarias, Vol. 36, No. 2, 2 de agosto de 2004.

Predicción de Solicitudes por Segundo HTTP para escalado proactivo de kubernetes: Un Enfoque Basado en Machine Learning y Series Temporales

Daniel Anunziata[1], Emilio Corti[1]
1 Laboratorio de Redes, Grupo de Optimización
Facultad de Ingeniería. Universidad Nacional de Río Cuarto
Ruta Nacional 36 km 601, CP-5800, Río Cuarto, Córdoba, Argentina
{danunziata , emiliocorti}@ing.unrc.edu.ar

Resumen

En esta línea de investigación, se realiza un análisis comparativo de diversos modelos predictivos para la tasa de solicitudes por segundo HTTP, con el objetivo de optimizar las decisiones de escalado proactivo en entornos Kubernetes. Se evalúan modelos individuales como Prophet y LSTM, así como híbridos (Prophet-LSTM, Prophet-GRU y ARIMA-LSTM) diseñados para capturar tanto la estacionalidad como las dependencias temporales presentes en las cargas de trabajo. La evaluación se lleva a cabo utilizando conjuntos de datos reales, como los logs de acceso web de una Copa Mundial y de la NASA [1][2], considerando métricas de rendimiento y el tiempo de entrenamiento y predicción. Los resultados preliminares sugieren que los modelos híbridos tienen el potencial de mejorar significativamente la precisión y eficiencia del autoescalado proactivo en comparación con los modelos individuales.

Palabras Claves: Kubernetes, autoescalado proactivo, machine learning, series temporales, optimización

Contexto

Los centros de datos, instalaciones críticas que albergan múltiples dispositivos de red, servidores y sistemas de climatización y respaldo eléctrico, se enfrentan a un aumento exponencial en la demanda de capacidad de cómputo y almacenamiento, lo que incrementa el consumo energético. Además, la escasez global de energía presenta la necesidad de mejorar la eficiencia del uso de recursos en estas instalaciones. Este contexto es la motivación de este trabajo, el cual forma parte de las actividades iniciales planteadas en el proyecto de investigación: “Gestión Eficiente del Despliegue de Servicios de Inteligencia Artificial en Cloud Computing”. Dicho proyecto integra al Programa “Desarrollo de Técnicas Inteligentes para Optimización de Ingeniería”, dependiente del Grupo de Optimización, perteneciente a la Facultad de Ingeniería de la UNRC.

Introducción

El autoescalado en Kubernetes es esencial para la gestión eficiente de recursos en entornos de nube dinámicos [3][4]. Los enfoques reactivos, como el Horizontal Pod Autoscaler (HPA), tienen limitaciones debido a la latencia en la respuesta a los cambios de carga [4][5]. Los modelos predictivos pueden anticipar la demanda, pero la precisión es clave. Existe una necesidad de modelos que capturen patrones complejos y se adapten a las variaciones de la carga de trabajo. Los modelos existentes a menudo carecen de la capacidad de capturar tanto la estacionalidad como las dependencias temporales [4][6].

Objetivos

Objetivo General: Evaluar modelos de predicción de solicitudes por segundo HTTP que faciliten el escalado proactivo en Kubernetes, integrando técnicas de Machine Learning y series temporales para optimizar el uso de recursos y mejorar la eficiencia operativa.

Objetivos Específicos:

- Analizar las ventajas y limitaciones de los enfoques reactivos tradicionales de autoescalado en Kubernetes.
- Identificar los patrones temporales y características específicas en las cargas de trabajo HTTP que pueden ser explotados por modelos predictivos.
- Comparar el rendimiento de los modelos propuestos utilizando métricas estándar (MSE, RMSE, MAE, R^2) y métricas específicas de Kubernetes (tiempo de respuesta a cambios de carga, eficiencia en el uso de recursos).
- Analizar el impacto del tiempo de entrenamiento y predicción en la viabilidad operativa de los modelos en entornos de producción.
- Desarrollar un prototipo de sistema de autoescalado proactivo basado en los modelos con mejor desempeño y validar su efectividad en un entorno de Kubernetes controlado.

Resultados Preliminares y Esperados

Los resultados preliminares indican que los modelos híbridos superan consistentemente a los modelos individuales en términos de precisión predictiva. En particular el modelo Prophet-LSTM muestra un rendimiento superior en datasets con patrones estacionales marcados, como el set de datos World Cup, con una reducción del RMSE de aproximadamente 15-20% respecto a los modelos individuales.

Se espera que la implementación de estos modelos en un sistema de autoescalado proactivo para Kubernetes permita:

- Reducir la latencia en la respuesta a cambios de carga respecto a enfoques reactivos.
- Mejorar la eficiencia en la utilización de recursos, minimizando tanto la subprovisión como la sobreprovisión.
- Proporcionar un marco adaptable que pueda ajustarse a diferentes patrones de carga y requisitos de aplicaciones.

Formación de Recursos Humanos

Esta investigación, que se enmarca dentro de un proyecto “Gestión Eficiente del Despliegue de Servicios de Inteligencia Artificial en Cloud Computing”, perteneciente a los Programas y Proyectos de Investigación Científica y Tecnológica (PPI) de la Universidad Nacional de Río Cuarto. Forma parte de las actividades del Laboratorio de Redes de la Facultad de Ingeniería, por lo que participan becarios y ayudantes alumnos, los cuales asisten en forma voluntaria fuera de su horario de clases. A su vez, es un espacio donde pueden realizar Prácticas Profesionales Supervisadas los estudiantes del último año de la carrera Ingeniería en Telecomunicaciones. Respecto la formación de posgrado, en la actualidad un maestrando se encuentra realizando su tésis con título “Infraestructura Dinámica con Observabilidad y Autoescalado para la Optimización de Recursos en Entornos Clusterizados”.

Referencias Bibliográficas

[1] 1998 World Cup Web Site Access Logs. Available at: https://ita.ee.lbl.gov/html/contrib/WorldCup.html

[2] 1995 NASA Web Site Access Logs. Available at: https://ita.ee.lbl.gov/html/contrib/NASA-HTTP.html

[3] Kubernetes Authors, "Horizontal Pod Autoscaler," Kubernetes Documentation, 2023.

[4] D. Guruge y C. Priyadarshana, "Hybrid Prophet-LSTM Models for Proactive Autoscaling in Kubernetes Environments," Journal of Cloud Computing: Advances, Systems and Applications, vol. 8, no. 2, pp. 123-145, 2025.

[5] J. Zhang, M. Wang y N. Guan, "Predictive Autoscaling for Microservice Applications: A Survey," IEEE Transactions on Cloud Computing, vol. 10, no. 3, pp. 1-15, 2023.

[6] P. C. Brebner, "Real-world performance modelling of enterprise service oriented architectures: delivering business value with complexity and constraints," ACM SIGSOFT Software Engineering Notes, vol. 36, no. 5, 2011

Proyecto: Sistema de Comunicaciones para Cube Sat

Leonardo Martí[1], Armando Lima[1], Gustavo Mercado[1]

[1]gridTICs – Grupo en Tecnologías de la Información y las Comunicaciones
Departamento de Electrónica / Facultad Regional Mendoza / UTN
teledat.radio@gmail.com, armando.f.lima@gmail.com, gmercado@frm.utn.edu.ar,

Resumen

El progreso experimentado en el ámbito de la tecnología aeroespacial a nivel global ha sido evidente en años recientes. Sin embargo, persisten desafíos en su aplicación a nivel local, particularmente en lo concerniente a la telemetría de satélites de reducidas dimensiones, tales como los denominados CubeSats. Uno de los principales obstáculos que se observan radica en la escasez de información accesible y detallada concerniente a la integración y operación de tales sistemas en proyectos desarrollados en el territorio argentino.

El propósito de este proyecto es sentar las bases para futuros desarrollos tecnológicos en el país, abordando tanto los aspectos técnicos como la necesidad de fomentar el crecimiento del sector aeroespacial nacional.Además, se busca aprovechar las oportunidades actuales de transporte espacial para diseñar y lanzar equipos satelizables, los cuales pueden albergar cargas útiles con objetivos científicos, educativos o comerciales. En todos los casos, se considera esencial contar con un módulo de telemetría que facilite la comunicación y supervisión del estado del satélite desde tierra.

Para ello, se plantea el desarrollo de un módulo de telemetría versátil, capaz de integrarse en diversos proyectos y garantizar una comunicación eficiente con estaciones terrestres, sin depender de las características específicas de las cargas útiles.Este módulo se fundamentará en circuitos integrados avanzados que combinarán en un solo chip funciones de radiofrecuencia, modulación y almacenamiento de datos, lo que permitirá explorar nuevas tecnologías de comunicación satelital.

El diseño del sistema se fundamenta en los protocolos de comunicación satelital convencionales, con el propósito de establecer una plataforma escalable y adaptable a aplicaciones futuras. Este estudio sentará las bases para el desarrollo de soluciones más avanzadas, contribuyendo así a la innovación y al fortalecimiento del sector aeroespacial en el país.

Palabras clave: Telemetría, CubeSats, Comunicaciones Satelitales, Electrónica Espacial, Protocolos de Comunicación, Desarrollo Aeroespacial Nacional

Contexto

El proyecto en cuestión se fundamenta en la cooperación del grupo de I+D gridTICs (UTN FRM) en el proyecto MendoSat MS-1. Este último es gestionado por el Polo Tics Mendoza, y desde el gridTICs se asume el compromiso de colaborar en diversos aspectos, entre los cuales se encuentra el proyecto de Sistema de Comunicaciones para Cube Sat.

El gridTICs se desempeña como stateholder, proporcionando recursos humanos especializados, incluyendo becarios, estudiantes. También contribuye con materiales de trabajo y un entorno de laboratorio experimental. Además, se propone fomentar trabajos finales en la carrera de

ingeniería electrónica en colaboración con la Cátedra de Proyecto Final de la UTN FRM

1. INTRODUCCIÓN

En los últimos años ha crecido el interés por el desarrollo y lanzamiento de pequeños satélites de órbita baja (LEO), incluidos los nanosatélites (de 1 a 10 kg) y los picosatélites (de 0,1 a 1 kg), especialmente las populares plataformas CubeSat.
El término CubeSat se utiliza para describir un pequeño satélite cuya forma unitaria básica es un cubo de 10 cm de arista, es decir, 1U. Las unidades CubeSat pueden juntarse para formar artefactos mayores, como los de 2U, 3U, 6U, etc. Los CubeSats deben cumplir los estándares definidos en la Especificación de Diseño de CubeSat, que incluye las directrices de seguridad en vuelo.

El desarrollo inicial de los CubeSats surgió con un propósito educativo La idea era proporcionar a los estudiantes experiencia práctica en actividades espaciales, permitiéndoles trabajar en el ciclo completo de un proyecto espacial, desde el concepto inicial hasta su operación en el espacio.

El subsistema de telemetría, seguimiento y control (TT&C) de un satélite establece una conexión entre éste y las instalaciones terrestres. La función TT&C garantiza el correcto funcionamiento del satélite. Como parte del bus de la nave espacial, el subsistema TT&C es necesario para todos los satélites, independientemente de su aplicación.

El desarrollo de un módulo de TT&C, busca crear un elemento común aplicable a distintos proyectos que requieran telemetría de control desde tierra, de manera independiente de las comunicaciones específicas de las cargas útiles. Para este proyecto, se propone utilizar circuitos integrados avanzados que implementan la sección de radiofrecuencia, la modulación y el almacenamiento en caché de datos en un solo chip. Estos circuitos son relativamente nuevos y hay poca información técnica disponible sobre ellos, por lo que este proyecto supone una oportunidad para profundizar en su conocimiento y en el uso de tecnologías modernas de manejo de sistemas de radio integrados. Además, el proyecto se basará en los protocolos de comunicación estándar empleados en las comunicaciones entre módulos internos de satélites.

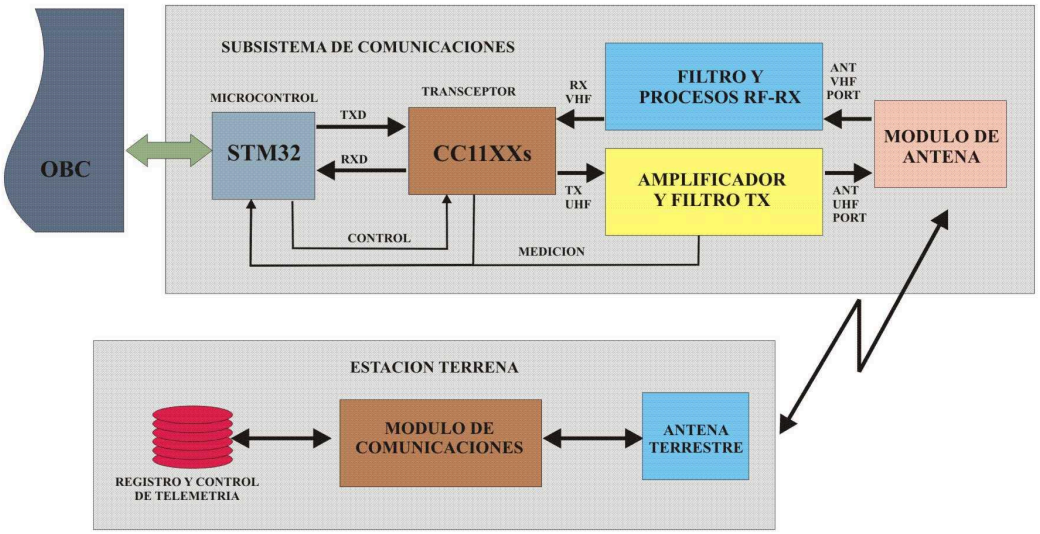
2. DESCRIPCIÓN DEL PROYECTO

Se propone el diseño de un sistema de Comunicaciones de Telemetría y Telecomando (TT&C) para un satélite de pequeño porte tipo CubeSat.

El sistema está compuesto por un transmisor y un receptor de RF embarcado que operan en bandas VHF y UHF, actuando como una interfaz entre el satélite y la estación terrestre. Desde la estación terrena se reciben comandos para operar el satélite, los cuales son enviados a la computadora de abordo (OBC), mientras que desde la OBC se reciben datos de telemetría (salud del satélite), los cuales son enviados a la estación terrena.Con los datos de telemetría recibidos en tierra se puede analizar el funcionamiento del satélite y prever acciones correctivas de ser necesarias.

El propósito de este estudio es proporcionar una base para el desarrollo futuro de proyectos espaciales, facilitando la implementación de niveles más elevados de complejidad y prestaciones más avanzadas.

Para la implementación del proyecto se emplearán tecnologías que posibiliten el desarrollo y montaje manual, garantizando al mismo tiempo el cumplimiento de los estándares de calidad satelital. En términos generales, el montaje en circuitos electrónicos para satélites se lleva a cabo manualmente, siguiendo normas específicas de montaje y soldadura.



Esquema TT&C

Figura 1: Diagrama esquemático de subsistema TT&C para small Leo satellite

2.1 Subsistema de Comunicaciones

En la figura 1, se exhibe un diagrama esquemático del subsistema, compuesto por los siguientes componentes: un módulo de procesamiento (microcontrolador), un módulo de transceptor, un módulo de filtros y amplificadores de RF, y un módulo de antena..

2.2 Módulo de procesamiento

Para el procesamiento de los datos recibidos desde la estación terrena a través del sistema de radio, se utilizará un microcontrolador que gestione el protocolo de datos y paquetes, y almacenará la información hasta que la OBC la requiera para su procesamiento.
En esta etapa, se utilizará un microcontrolador de la línea STM32 por su versatilidad, con el objetivo de emplear posteriormente un microcontrolador similar, pero más robusto y diseñado para ambientes hostiles.

2.3 Módulo de transceptor

El **CC112X** pertenece a una familia de transceptores RF de alto rendimiento y bajo consumo, diseñados para operar junto con un microcontrolador asociado.
Automatiza todas las tareas comunes relacionadas con RF, aliviando significativamente la carga de trabajo del sistema de control.
En su arquitectura en bloques, el **CC112X** se divide en dos secciones principales: una dedicada a la funcionalidad **RF** (generación de frecuencias, modulación, control de niveles, etc.) y otra destinada al soporte digital (conversión a digital, demodulación, memoria de datos, etc.).

2.4 Módulo de filtros y amplificadores de RF

Si bien el uso del CC112X como transceptor simplifica significativamente el diseño, en este caso es fundamental prestar especial atención a los aspectos de radiofrecuencia. El equipo estará expuesto a diversas señales terrestres que pueden generar interferencias, por lo que la recuperación limpia de las señales de control enviadas desde la estación

terrena es crucial, ya que constituyen el canal principal de comunicación con el satélite.
Para garantizar un desempeño óptimo del receptor, se diseñará un sistema de filtros y amplificadores de entrada que maximicen su eficiencia.
Además, es importante considerar que el equipo debe recibir mientras transmite en otra banda, lo que supone un desafío adicional. La potencia de transmisión puede superar con creces los niveles máximos tolerables por el receptor en condiciones normales de operación, por lo que será necesario implementar medidas para mitigar este impacto.

2.5 Módulo de antena

En este trabajo, desarrollaremos un sistema básico de antenas similares a las utilizadas a bordo, conectadas directamente a los puertos de la placa de radio. En la práctica, estos puertos deberán vincularse a un módulo encargado del despliegue de las antenas en vuelo, dado que, durante el transporte y lanzamiento, el satélite completo está sujeto a restricciones de espacio.
Los ensayos comenzarán con antenas tipo dipolo y monopolo, dispuestas en una estructura de prueba del satélite, con el objetivo de estudiar de manera aproximada su rendimiento, patrón de radiación, ganancia y otros parámetros relevantes.

3. DESARROLLO DEL PROYECTO

3.1 Firmware

El firmware estará integrado en un microcontrolador STM32. Este dispositivo es conocido por su alta eficiencia, su amplio ecosistema de desarrollo y su soporte para aplicaciones industriales y aeroespaciales. El firmware será una aplicación básica de recepción de paquetes, almacenamiento y transmisión a petición de parámetros. El enfoque del proyecto está en la sección de RF, ya que se documentará debidamente para que pueda evolucionar hasta alcanzar una versión final. Sistema de comunicaciones - RF

3.2 Entorno de desarrollo y Lenguaje de Programación

Se priorizará el uso de software basado en código abierto, compatible con sistemas operativos basados en Linux. El lenguaje de programación de los microcontroladores será C, orientado a microcontroladores.
La implementación se realizará a través de entornos de desarrollo previamente evaluados. La elección dependerá de lo más habitual en el ámbito del desarrollo satelital. Entre los posibles candidatos se encuentran:

- Visual Studio Code (VSC) con extensiones para microcontroladores.
- STM32CubeIDE el entorno oficial para microcontroladores STM32.

3.2 Diseño Físico

Se emplerá tecnología de componentes de montaje superficial (SMD). El diseño estará orientado a utilizar placas de circuito impreso (PCB) de doble faz, que sean fácilmente fabricables a nivel local, fomentando la producción nacional y la accesibilidad para quienes deseen experimentar con el proyecto.

3.3 Metodología de Desarrollo

El desarrollo del equipo se realizará en módulos independientes, de modo de efectuar la validación y prueba del funcionamiento de cada módulo de manera individual, facilitando así su integración en etapas posteriores. Luego, se diseñará un circuito impreso compatible con entornos satelitales, adecuado para su implementación en un CubeSat estándar.
Este enfoque modular, basado en tecnologías accesibles, busca combinar los estándares de calidad aeroespacial con la promoción del desarrollo tecnológico local.

4. OBJETIVOS

4.1 Objetivo Principal

El objetivo principal del proyecto es establecer las bases del diseño de un sistema de telemetría y telecomando operativo.
El proyecto consiste en construir, probar y documentar los comandos y modos operativos de los módulos transceptores que se van a utilizar, así como los circuitos de radiofrecuencia asociados hasta el sistema de transporte de señal a la antena.

4.2 Objetivos Específicos

1. Estudiar y diseñar modelo conceptual del sistema de comunicaciones embarcado.
2. Diseñar e implementar prototipo funcional de módulo de Transceptor para configuración y programación.
3. Desarrollar la etapa de radiofrecuencia y los circuitos de entrada/salida de los módulos transceptores.
4. Seleccionar el microcontrolador y la plataforma de programación, estableciendo normas y analizando software existente.
5. Diseñar e implementar módulo GSE (ground support equipment) de OBC y Estación Terrena
6. Implementar módulos de comunicación, verificar configuraciones y desarrollar la lógica de telemetría.
7. Definir parámetros de control y efectuar ensayos de RF.
8. Ensamblar los subsistemas en modo FlatSat para pruebas de integración y corrección de fallos.
9. Implementar una red de pruebas para evaluar la velocidad de comunicación y los límites del enlace de radio.
10. Validar el funcionamiento del software, los componentes y el montaje del sistema.
11. Diseñar el circuito final incorporando mejoras tras las pruebas.
12. Diseñar PCB basada en el esquemático final.
13. Implementar modelo de ingeniería y realizar pruebas iniciales del prototipo satelizable.
14. Medir y validar parámetros requeridos en el modelo de ingeniería.

5. FORMACIÓN DE RECURSOS HUMANOS

Uno de los principales objetivos del proyecto es la capacitación de los recursos humanos.
La meta es fortalecer la capacidad para realizar investigación científica, generar conocimientos y facilitar la transferencia de tecnología que permita el desarrollo humano.
Para lograr estos objetivos se dispuso del siguiente personal:
Un Investigador principal
Un Investigador de apoyo y aspirante a doctorado
Dos Becarios alumnos (Beca BIS)
Un Tesista de carrera de grado

6. BIBLIOGRAFÍA

(Weston et al 2023) Sasha Weston, “Small Spacecraft Systems”, Ames Research Center, NASA, January 2023

(Cappelletti et all 2021) Cappelletti C, Battistini S, Malphrus B, “CubeSat Handbook: From Mission Design to Operations”, Academic Press Elsevier, 2021

(Villela 2019) Thyrso Villela, Cesar A. Costa, Alessandra M. Brandão, Fernando T. Bueno, Rodrigo Leonardi, “Towards the Thousandth CubeSat: A Statistical Overview”, doi.org/10.1155/2019/5063145

(Guest Arthur 2017) Norman Guest “Telemetry, Tracking, and Command (TT&C)” in Handbook of Satellite Applications, Springer 2017

(Giunta 2023) A. Giunta, L. Muñoz, E. Chediack, “MS 1 Requerimientos Generales”, en PT MS1 Proyecto sistema de conectividad satelital para IOT (MendoSat-I), Doc Nr MS1-0GG-GREQ-01-A, Polo Ticas Mendoza, Ago 2023

(Huertas 1988) Carlos Huertas, "Satélites 1", Editorial Hasa, 1988.

(Huertas 1990) Carlos Huertas, "Satélites 2", Editorial Hasa, 1990.

(Texas Instruments 2016) Texas Instruments, *CC1125 High-Performance RF Transceiver*, Datasheet, 2016. Disponible en: https://www.ti.com/lit/ds/symlink/cc1125.pdf

Tecnologías de IoT y visión artificial con aplicaciones medioambientales y robótica

Jorge Osio[1,2], Juan Salvatore[1], Mauro Salina[1] , Marcelo Cappelletti[1,2] , Diego Montezanti[1,3], Juan Piñeiro[1], Braian Pezet[1], Lucia Osés[1], Martín Paez[1],Emanuel Julio[1], Matías Busum Fradera[1], Facundo Chazarreta[1], Lucas Olivera[1,4], María Florencia Ayala[1], Diego Encinas[1,3], Martín Morales[1,5]

[1] Programa TICAPPS, Univ. Nac. Arturo Jauretche, Florencio Varela (1888), Argentina.
[2] Grupo de Control Aplicado, Instituto LEICI (UNLP-CONICET), La Plata (1900), Argentina.
[3] Instituto de Investigación en Informática LIDI (III-LIDI) - Facultad de Informática - UNLP – Centro Asociado CIC.
[4] Becario de la Comisión de Investigaciones Científicas, Pcia. Bs. As. (CICPBA), La Plata (1900), Argentina.
[5] Centro UTN CODAPLI-FRLP, La Plata (1900), Argentina.
{josio, mcappelletti, dmontezanti, jsalvatore, mdsalina, dencinas, martin.morales}@unaj.edu.ar

Resumen

El presente proyecto se enfoca en líneas de investigación que involucran nuevas tecnologías como Internet de las Cosas (IoT), visión artificial y robótica. La investigación se centra especialmente en las técnicas de Smart IoT, las técnicas de aprendizaje automático aplicadas sobre dispositivos de bajas prestaciones de procesamiento y el aprendizaje profundo para el reconocimiento de objetos en aplicaciones de visión por computadora. Se espera que los resultados obtenidos permitan proporcionar soluciones a problemáticas socio-productivas presentes en la región, como son el cuidado del medioambiente, la educación y la innovación en el sector productivo mediante tecnologías de la industria 4.0, en la evolución hacia la industria 5.0.

Entre los temas de investigación que se desarrollan en esta línea, se incluyen el diseño e implementación de técnicas de visión por computadora, con los objetivos de agregar funcionalidades a dispositivos robóticos (como la clasificación de residuos) y el análisis para descontaminación de suelos en aplicaciones agrícolas, entre otros. Además, mediante el agregado de control y supervisión remota (utilizando herramientas de IoT), se busca proporcionar a estos dispositivos la autonomía suficiente para la realización de tareas en entornos abiertos.

En el área de Smart IoT, se ha desarrollado una plataforma IoT que incluye los servicios esenciales, como un *broker* MQTT, una base de datos, un servicio de gráfica de datos y una plataforma de configuración orientada a flujos (NodeRed) [1]. En esta plataforma se está desarrollando un servicio de análisis de datos que utiliza aprendizaje automático para predicción y toma de decisiones. Por último, actualmente se están investigando aspectos de protección de la información en sistemas IoT.

Palabras clave: *Aprendizaje automático, procesamiento de imágenes, Smart IoT, Visión por computadora.*

Contexto

Las líneas de I/D presentadas en este trabajo son desarrolladas por el grupo IoT&IA y están incluidas dentro del Programa TICAPPS (TIC con aplicaciones de interés social) de la Universidad Nacional Arturo Jauretche (UNAJ), Resolución N° 064/17, bajo la dirección del Dr. Ing. Martín Morales.

1. Introducción

Las tecnologías de Internet de las Cosas (IoT) son necesarias para el envío y almacenamiento de grandes cantidades de datos en la nube [2]. Por otro lado, la visión por computadora

permite la detección de características y patrones, y la adquisición de información que puede utilizarse para manipular objetos, automatizando acciones y procesos mediante el guiado autónomo, entre otras posibilidades.
En los últimos años las tecnologías de IoT e inteligencia artificial (IA) han evolucionado notablemente, posibilitando la innovación tecnológica en todas las áreas en las que se aplican.
Por otra parte, el avance de las tecnologías de sistemas embebidos está haciendo posible que muchos de los algoritmos de aprendizaje automático se puedan ejecutar localmente en las capas de *Edge* o *Fog Computing*, cercanas a los nodos sensores. Esto resulta alentador en una gran variedad de aplicaciones de Smart IoT, ya que permite que el sistema pueda seguir operando de forma autónoma cuando se pierde conexión con el servidor [3]. Los dispositivos de estas capas son capaces de almacenar y analizar sus propios datos, de manera que las decisiones se pueden tomar localmente proveyendo alta disponibilidad y la resiliencia.
Dentro de la visión por computadora, la identificación de objetos es uno de los temas con mayor aplicación en diferentes ámbitos. Básicamente, diferentes algoritmos son entrenados para buscar patrones que se correspondan con un objeto en particular. Una de las técnicas más utilizadas, conocida como *Deep Learning* [6], consiste en aprender a encontrar características de forma automatizada en las diferentes capas del algoritmo, lo que *a posteriori* posibilita la toma de decisiones correctas y la considerable reducción en la cantidad de falsos positivos en el análisis. *Deep Learning* es una evolución del Aprendizaje Automático [6] y ha posibilitado la detección de características y clasificación de imágenes de forma mucho más eficiente junto con las innovaciones en el manejo de datos a gran escala y la amplia variedad de servicios que existen para el análisis y procesamiento de información en la nube (*Cloud Computing*). Resumiendo, las técnicas de Deep Learning (Aprendizaje Profundo) proveen soporte para el diseño de aplicaciones de visión por computadora, por lo que son la base de las investigaciones sobre detección de características en imágenes [6].
Po su parte la Internet de las Cosas Inteligentes (SIoT) promete importantes avances en diferentes áreas de aplicación. Las aplicaciones de SIoT utilizan un conjunto de protocolos de comunicación e interfaces que pueden ser implementados mediante herramientas propias de los sistemas embebidos, como las desarrolladas en [4-5]. Se espera que los resultados que se obtengan a partir de las investigaciones sobre Smart IoT se puedan aplicar en el monitoreo y la automatización de procesos, y en el almacenamiento masivo de datos provenientes de nodos sensores (Big Data).
Por último, la importancia de disponer de un servidor propio, en el que es posible desplegar servicios para IoT e inteligencia artificial, posibilita no sólo obtener grandes cantidades de datos, sino también realizar análisis y procesamiento sin depender de servicios externos. De esta manera, es posible implementar aplicaciones de forma personalizada, configurando acciones de automatización y control de manera eficiente en un mismo sistema [1]. En esta instancia, se torna relevante considerar los aspectos de seguridad en las plataformas IoT; debido a ello, se busca agregar un cifrado a los datos de extremo a extremo (desde los nodos hasta la base de datos del servidor).

Aplicaciones y desarrollos en curso

Entre las aplicaciones que se encuentran en desarrollo, combinando técnicas de IoT, procesamiento de datos e imágenes y *Machine Learning* se destacan:

1) Detección automática de objetos reciclables.
2) Detección de gestos para el control de dispositivos por parte de personas con problemas motrices.
3) Medición de la calidad del aire, monitoreo de la calidad del agua y obtención de magnitudes climatológicas
4) Visión artificial aplicada a robótica [5].

El reciclado de desechos es de suma importancia para el cuidado del medioambiente, debido a que posibilita la reutilización de elementos u objetos que de otro modo serían desechados y contribuirían al incremento de la basura y al daño ambiental permanente. Sin embargo, el problema radica especialmente en el esfuerzo requerido para clasificar y separar los residuos inorgánicos. Debido a esto, en la línea de investigación se está desarrollando un sistema basado en redes neuronales artificiales que permite detectar y clasificar objetos reciclables para su posterior separación [7].
Respecto a la visión por computadora, se está llevando adelante la implementación de un sistema robótico que sea capaz de detectar obstáculos y recorrer espacios cerrados, mediante procesamiento de imágenes, para dotarlo con capacidades autónomas y utilizarlo en aplicaciones de calidad de aire y desinfección de suelo [8]. Además, se está implementando un brazo robótico que permite separar objetos previamente identificados mediante visión artificial [8].
Por último, se está desarrollando una aplicación de SIoT, que obtiene imágenes provenientes de un sistema embebido, las procesa mediante un algoritmo de detección de gestos faciales y permite accionar diferentes dispositivos por cada gesto detectado, generando así acciones sin la necesidad de emisión de ningún sonido ni acción física. Esto podría resultar de gran utilidad para personas con problemas motrices y dificultades para el habla [9].

2. Líneas de Investigación y Desarrollo

El grupo de investigación que se ha constituido en la UNAJ es multidisciplinario, y sus miembros cuentan con experiencia en sistemas de cómputo de altas prestaciones, arquitecturas distribuidas, resiliencia, sistemas embebidos, IoT, aprendizaje automático y robótica.
Una de las líneas de I/D que se presentan en este trabajo está basada en el estudio y desarrollo de herramientas alternativas para el procesamiento de imágenes a partir de algoritmos de aprendizaje automático y visión por computadora. La línea referente a sistemas SIoT pretende, en primera instancia, evaluar las herramientas tecnológicas disponibles para determinar las posibles aplicaciones, pero también sus limitaciones.
En cuanto a la línea vinculada con la robótica [8], se encuentra en desarrollo la búsqueda de aplicaciones innovadoras, como la separación automática de objetos reciclables, la capacidad de detección de gestos y la identificación de distintos entornos por parte de robots móviles.

Temas de Estudio e Investigación

Los datos recopilados a través de robots y nodos sensores en diferentes aplicaciones pueden resultar en grandes cantidades de información. Si bien esto requiere de involucrar técnicas de Big Data para su tratamiento, también permite obtener una mayor precisión en las predicciones de los sistemas de *Machine Learning*, ya que los modelos se alimentan con mayor volumen de datos útiles para su aprendizaje.
Los aportes originales no sólo tienen que ver con el manejo de la información, sino también con las formas innovadoras en las que se puede aplicar el aprendizaje automático en visión por computadora y robótica. El ejemplo más claro consiste en la posibilidad de disponer de un sistema de detección de objetos reciclables para la separación automática de residuos. Por otra parte, la detección automática de gestos puede ser una alternativa viable de comunicación para personas con problemas motrices y del habla.
Por último, respecto a los aspectos de protección de la información en sistemas IoT, se han implementado comunicaciones seguras basadas en SSL que permiten el encriptado entre broker y gateway. Para la seguridad de los dispositivos finales y la comunicación con el servidor; la integridad (algoritmo de Hash), confidencialidad (encriptado simétrico AES) y los métodos de autenticación para el acceso remoto a la información.

Resultados alcanzados:

Entre las mejoras asociadas a los métodos de detección de objetos reciclables, se ha logrado incrementar cinco veces el tamaño del dataset original, alcanzando una cantidad de 15000 imágenes. También, se perfeccionaron los algoritmos para identificar y clasificar varios objetos en una misma imagen, en tiempo real. Finalmente, se optimizó el algoritmo usando la última versión de YOLO para que funcione correctamente en sistemas de cómputo de bajas prestaciones (como por ejemplo la Raspberry Pi 3 Model B+ y dispositivos móviles) sobre los que se ejecutaron las aplicaciones, mediante el uso de la arquitectura MobileNet [7]. Gracias a estas contribuciones, se pudieron mejorar notablemente los resultados de la clasificación, alcanzando muy altos porcentajes de certeza en la detección (del orden de 0.94, es decir una probabilidad de acierto muy cercana a 1). Por su parte se observó una particularidad en las clases vidrio (glass) y plástico (plastic) pudieron alcanzar porcentajes de certeza por debajo de los alcanzados por el resto de los materiales. Esto se debe a que comparten características similares, como la transparencia.

En el desarrollo del sistema SIoT capaz de detectar gestos para el accionamiento de dispositivos, se utilizó un esp32-CAM para la obtención de imágenes en conjunto con una librería muy eficiente llamada Mediapipe, a la que se le pueden ajustar los parámetros para obtener una detección más precisa. Gracias a la comunicación a través del protocolo MQTT, entre diferentes dispositivos y aplicaciones conectadas al bróker, es posible realizar acciones en base a los gestos detectados en tiempo real. El sistema desarrollado, basado en IoT y visión por computadora, demostró ser efectivo en la detección de gestos faciales bajo condiciones controladas, permitiendo a usuarios con movilidad reducida interactuar con su entorno mediante gestos, como, por ejemplo, controlar dispositivos domésticos o enviar notificaciones. Las pruebas validaron la precisión en la detección de gestos como guiños, apertura de boca y muecas, con una latencia aceptable en tiempo real gracias al uso de algoritmos como MediaPipe y la integración eficiente del protocolo MQTT. Sin embargo, se identificaron limitaciones en escenarios con iluminación variable, uso de accesorios (ej. lentes) o ángulos pronunciados (>30°), donde la precisión disminuyó debido a errores en la generación de la malla facial [9].

Por su parte, entre las aplicaciones IoT desarrolladas, se están obteniendo datos de las sondas para medición de parámetros en arroyos y de cultivos en huertas. Esta información permitirá detectar el vertido de desechos en arroyos y mejorar las técnicas para optimizar la cosecha, respectivamente. En esta línea, se están migrando los servicios a la nueva plataforma IoT desarrollada, junto con las mejoras en el HW respecto a la disminución del consumo energético y el incremento de la autonomía de la alimentación a batería con el agregado de paneles solares.

Objetivos esperados:

- Promover la generación de conocimiento y el desarrollo de instrumentos que puedan aplicarse en la formación de recursos humanos para investigación.
- Fortalecer la actividad de investigación y vinculación que contribuya a remediar las problemáticas del territorio de influencia de la UNAJ.
- Diseñar y optimizar diferentes arquitecturas y redes para IoT, en función de la aplicación.
- Evaluar los modelos de visión artificial obtenidos para sistemas embebidos (con bajas prestaciones de procesamiento y bajo consumo energético). Reemplazar la tecnología ESP32 por una raspberry para la detección de objetos y el control del brazo robótico.
- Implementar un sistema de calibración automático que ajuste dinámicamente los umbrales de detección y la incorporación de técnicas avanzadas de aprendizaje profundo (como redes neuronales convolucionales y algoritmos de segmentación semántica) para optimizar el procesamiento de imágenes en la detección de gestos. Asimismo, se evaluarán soluciones eficientes que no dependan exclusivamente de hardware especializado, explorando optimizaciones en modelos de

aprendizaje automático, como la cuantización o poda de redes, orientadas a reducir la latencia y el consumo de recursos en dispositivos de Edge Computing.
- Respecto a la seguridad en IoT, se evaluará la integración de mecanismos de seguridad robustos (TLS/DTLS) y el despliegue de procesamiento distribuido en microcontroladores para reducir la dependencia del servidor central.

Formación de Recursos Humanos

Uno de los principales objetivos del Programa TICAPPS, dentro de la temática de las líneas de I/D presentadas en este trabajo, es la formación de recursos humanos, tanto de docentes investigadores como de estudiantes.
Dentro de la temática de la línea de I+D, todos los miembros del proyecto participan en el dictado de asignaturas de la carrera de Ingeniería Informática de la UNAJ.
En este proyecto existe cooperación a nivel Nacional. Hay tres Doctores, dos Magisters, un Especialista, cuatro Doctorandos y un integrante realizando su Maestría en temas relacionados.
Además, hay 4 becarios y estudiantes avanzados realizando las PPS de final de carrera en la temática. Actualmente, se encuentran en curso tres Tesis de Doctorado y tres becas de Estímulo a las Vocaciones Científicas del Consejo Interuniversitario Nacional (Becas EVC – CIN) y 2 becas UNAJ de investigación relacionadas directamente con las líneas de I/D presentadas.

3. Bibliografía

[1] F. Chazarreta, M. Busum Fradera, J. R. Osio, J. E. Salvatore, Morales D. M. “Diseño de una plataforma de servicios IoT para aplicaciones de interés social”, IX CONAIISI, 2022.

[2]L. Fatmasari Rahman,” Choosing your IoT Programming Framework: Architectural Aspects”, 2016 IEEE 4th International Conference on Future Internet of Things and Cloud

[3] Medina, S., Montezanti, D., Gómez D’Orazio, L., Garay, F., De Giusti, A., & Naiouf, M. (2023, June). Distributed Architectures Based on Edge Computing, Fog Computing and End Devices: A Conceptual Review Incorporating Resilience Aspects. In Conference on Cloud Computing, Big Data & Emerging Topics (pp. 31-44). Cham: Springer Nature Switzerland

[4] J. Osio, J. Salvatore, M. Salina, M. Cappelletti, D. Montezanti, L. Olivera, M. Paez, F. Palacio, M. Busum Fradera, F. Chazarreta, “Tecnologías de Smart IoT para la solución de problemas en el medio socio productivo”, UNNOBA, Ciudad de Junín, WICC 2023.

[5] J. Osio, J. Salvatore, M. Salina, M. Cappelletti, D. Montezanti, L. Olivera, M. Paez, F. Palacio, M. Busum Fradera, F. Chazarreta, “Nuevas tecnologías de IoT y visión artificial para el diseño de sistemas con aplicaciones de interés social”, UNP, Puerto Madryn, WICC 2024.

[6] N. Buduma, “Fundamentals of Deep Learning”, Editorial O’reilly. 2017

[7] L. Osés, B. Pezet, M Salina, JR Osio, MA Cappelletti, M Morales. Técnicas de Deep Learning aplicadas a un sistema de clasificación de objetos para un recolector de residuos inteligente. 52 Jornadas Argentinas de Informática Congreso Argentino de Ciencias de la Computación (JAIIO). 2023

[8] M. Jamshidi y P. J. Eicker, Robotics and remote system for hazardous environments, Prentice Hall, New Jersey, Estados Unidos (1993)

[9] M. Busum, F. Chazarreta, J. Salvatore, J. Osio, M. Paez, M. Morales. Desarrollo de técnicas de Smart IoT aplicadas a la detección de gestos para el accionamiento de dispositivos. 11° Congreso Nacional de Ingenierías en Informática y sistemas de Información (CoNaIISI 2023)

USO DE TECNOLOGÍA DE BLOCKCHAIN PARA MANTENER LA INTEGRIDAD DE LOS DATOS GENERADOS POR SENSORES EN APLICACIONES DE IOT

Cairo, Lucía[1], Corral, Rocío[1], Sanchez Lanza, Agustina[1], Ernesto Chediak[1] ,Godoy Pablo Daniel[1,2,3], Fontana Daniel[1], Marianetti Osvaldo[1]

[1]Facultad de Ingeniería, Universidad Nacional de Cuyo.

[2]Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Cuyo.

[3] ITIC, Universidad Nacional de Cuyo.

osvaldo.marianetti@ingenieria.uncuyo.edu.ar

RESUMEN

En el presente trabajo se analiza el aporte de la tecnología de Blockchain al objetivo de obtener un sistema descentralizado y distribuido en una aplicación de Internet de las Cosas. Esta implementación permite intercambio de datos a prueba de manipulaciones, de manera descentralizada y mejorando las condiciones de escalabilidad y la latencia de la aplicación. Se presenta el desarrollo de un prototipo útil para el diseño de sistemas de control automatizado en el contexto de entornos de IoT, haciendo uso de blockchain para garantizar la confiabilidad y la integridad de los datos sin incrementar significativamente la complejidad del firmware.

Palabras clave: blockchain, IoT, seguridad, sensores.

Área: ARSO (Arquitectura, Redes y Sistemas Operativos).

CONTEXTO

El presente artículo describe algunas de las experiencias desarrolladas por estudiantes y docentes de la carrera de Licenciatura en Ciencias de la Computación de la Facultad de Ingeniería de la UNCUYO, en el marco de un proyecto de la convocatoria "Proyectos SIIP 2025-2027" de dicha universidad.

Este trabajo es un caso de estudio particular. En dicho caso se considera la necesidad de mejorar la eficiencia energética y la seguridad en el hogar. Este requerimiento ha impulsado el desarrollo de sistemas inteligentes de control de dispositivos eléctricos, debido a la naturaleza de su funcionamiento y el riesgo que conlleva. En sistemas de calefacción tradicionales, el ajuste manual a menudo genera un uso ineficiente de energía, además de aumentar los riesgos al olvidar apagar el dispositivo en momentos innecesarios. Ante esta problemática, el uso de Internet de las Cosas (IoT) permite desarrollar sistemas automatizados que no solo controlen estos dispositivos de forma autónoma, sino que además faciliten el monitoreo y control remoto.

La integración de sensores ambientales en sistemas de IoT ha avanzado considerablemente, permitiendo medir variables clave como temperatura, humedad y calidad del aire. Estos sensores, en combinación con microcontroladores económicos pero prácticos como el ESP32, posibilitan el desarrollo de sistemas de control precisos, adaptables a diversos entornos y capaces de optimizar el uso de dispositivos eléctricos.

Asimismo, la tecnología blockchain, reconocida principalmente por su aplicación en criptomonedas, ha encontrado una utilidad emergente en el control y trazabilidad de datos en sistemas de IoT. Gracias a su estructura distribuida e inmutable, blockchain puede garantizar la integridad y trazabilidad de la información almacenada, lo que permite confiar en que los datos y comandos del

sistema no sean manipulados de forma indebida. Al combinar IoT con blockchain, es posible asegurar que los dispositivos respondan a condiciones predefinidas de manera segura, y que toda acción quede registrada para su verificación posterior.

A su vez, es posible encontrar múltiples herramientas low-code para facilitar la integración de sistemas de monitoreo y control web con dispositivos IoT sin necesidad de escribir grandes cantidades de código de bajo nivel para este fin.

Este proyecto integra IoT y una plataforma cloud para crear un sistema de control de estufa eléctrica que ofrece autonomía y seguridad.

1. INTRODUCCIÓN

1.1 Objetivos

Como se mencionó anteriormente, el objetivo principal del trabajo es desarrollar un sistema de control automatizado y remoto para una estufa eléctrica, usando un microcontrolador ESP32 junto con un sensor de calidad del aire y temperatura (BME680). La finalidad es mantener la temperatura ambiente estable y segura en espacios interiores (principalmente en una habitación de un hogar), activando la estufa cuando la temperatura desciende por debajo de un umbral predefinido.

En este contexto, el sistema creado no solo permite la automatización del encendido y apagado de la estufa mediante un relé, sino que también habilita su control remoto a través de una aplicación de la plataforma Blynk. A través de esta aplicación, el usuario puede monitorear en tiempo real los niveles de temperatura y calidad del aire y también puede ajustar el funcionamiento de la estufa desde cualquier lugar con acceso a Internet. Esto último, permite una solución cómoda para el usuario, maximizando la eficiencia energética y disminuyendo riesgos con el control manual en escenarios inesperados o atípicos.

A su vez, el proyecto contempla la integración de indicadores LED para brindar retroalimentación visual sobre el estado de la conexión WiFi y el funcionamiento de la estufa.

En resumen, los objetivos específicos son:

- Automatizar el encendido y apagado de la estufa basado en la temperatura ambiente medida por el sensor BME680.
- Evaluar continuamente la temperatura y la calidad del aire.
- Permitir al usuario el control y monitoreo remoto del sistema mediante la plataforma Blynk.
- Ofrecer un sistema confiable y seguro mediante el uso de indicadores visuales.

1.2 Trabajos Relacionados

La solución propuesta consiste en un sistema embebido, cuya integración se explica en esta sección. En la Figura 1 se puede visualizar la arquitectura de componentes de la solución y se detalla la integración de componentes de hardware y software que permiten medir, comunicar y responder a condiciones del entorno de forma dinámica para la implementación de todo el sistema.

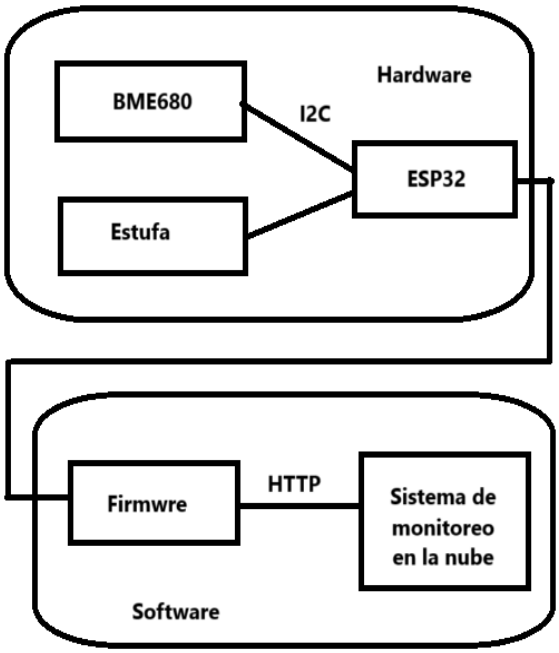


Figura 1: Arquitectura del sistema

El componente central de hardware de procesamiento del sistema es una placa ESP32, que tiene la capacidad de manejar las lecturas que provee el sensor ambiental BME680, controlar el dispositivo externo (la estufa) a través de un relé, conectarse a la red WiFi y enviar datos a la plataforma externa en la nube. Todo esto conforma el sistema que permite la supervisión y el control de la estufa. El firmware es la pieza de software que permite el control y funcionamiento del sistema a nivel de hardware, permitiendo que el ESP32 gestione tanto los sensores como los actuadores, la comunicación WiFi y la interacción con el usuario a través de la aplicación cloud mediante el protocolo HTTP. Está desarrollado en C usando el entorno Arduino IDE, y su diseño modular facilita el control sobre cada componente del sistema.

1.3 Estado actual de avance del proyecto

Se han desarrollado las siguientes funciones:

a. Inicialización y Configuración del Sistema.
b. Conexión a la red WiFi: Utilizando las credenciales correspondientes, el ESP32 intenta conectarse a la red WiFi. Esta conexión es esencial para que el sistema pueda enviar datos a la aplicación cloud y recibir comandos del usuario de forma remota.
c. Inicialización del Sensor: Haciendo uso de la biblioteca Zanshin_BME680.h.
d. Inicializar la interfaz de comunicación I2C, que permite la integración con el ESP32.
e. Conexión con Blynk: La función Blynk.begin() inicializa la conexión con la plataforma en la nube Blynk. .
f. Configuración de Temporizadores (Timers).
g. Gestión de Sensores
h. Lectura de Sensores
i. Control de Temperatura
j. Control de Actuadores (Relé y LEDs)
k. Control del Relé
l. Indicadores LED s.
m. Comunicación con la Plataforma Blynk. El sistema también permite la interacción del usuario a través de la plataforma Blynk, permitiendo el control remoto y la visualización de datos en tiempo real desde cualquier dispositivo conectado a internet. En el firmware se configuraron las siguientes funciones de control:
n. Envío de Datos
o. Recepción de Comandos.

El software del sistema complementa el hardware al proporcionar una interfaz gráfica y comunicación en tiempo real. Para este trabajo se implementaron dos alternativas compatibles para este sistema con distinta tecnología y características. La primera implementación utiliza la aplicación Blynk, una plataforma IoT cloud low-code enfocada en la experiencia de usuario. La segunda implementación utiliza blockchain como una base de datos distribuida y confiable junto con un contrato inteligente para guardar el estado del sistema y accionar actuadores. Ambos sistemas desempeñan la misma función y son independientes entre sí.

2. LINEAS DE INVESTIGACION y DESARROLLO

2. Integración de blockchain y contratos inteligentes con internet de las cosas.

Edge Computing consiste en procesar los datos lo más cerca posible del lugar donde se generan (en el dispositivo o sensor). Esto reduce la latencia y el uso de ancho de banda, y mejora la seguridad. Fog Computing extiende el concepto de Edge Computing, distribuyendo el procesamiento a lo largo de una red jerárquica, donde algunos datos se

procesan en nodos intermedios antes de llegar al servidor central. Estas soluciones buscan mejorar la eficiencia, escalabilidad y seguridad en comparación con las arquitecturas centralizadas tradicionales.

-Hipótesis de Trabajo

El uso de plataformas de Blockchain para el procesamiento distribuido en aplicaciones de IoT puede reducir la dependencia de arquitecturas centralizadas y minimiza los riesgos asociados con la manipulación o pérdida de información, todo ello sin comprometer la escalabilidad ni la eficiencia del sistema.

-Objetivos

1. Descentralización del procesamiento de datos. Eliminar la dependencia de servidores centrales: Usar Blockchain para coordinar y gestionar la distribución de tareas de procesamiento entre nodos de IoT sin depender de una autoridad centralizada.
2. Distribución de la carga de procesamiento mediante contratos inteligentes: Automatizar la asignación de tareas de procesamiento de datos utilizando contratos inteligentes, asegurando que los dispositivos más cercanos o capaces procesen los datos de manera eficiente.
3. Redundancia descentralizada: Crear copias redundantes de los datos a través de la red Blockchain, garantizando que los datos no se pierdan en caso de fallos en los nodos.
4. Optimización del uso de recursos y escalabilidad. Asignación dinámica de recursos: Utilizar Blockchain para coordinar de manera eficiente el procesamiento distribuido, asignando las tareas a los nodos de IoT en función de su capacidad y cercanía a la fuente de datos.
5. Reducir la latencia y mejorar la eficiencia. Minimizar la latencia en la verificación de datos.

3. RESULTADOS OBTENIDOS/ESPERADOS

3. Prototipo de un sistema de control automatizado en IoT, haciendo uso de blockchain para garantizar la confiabilidad y robustez de los datos.

Se ha desarrollado un prototipo de sistemas de control automatizado en IoT, haciendo uso de blockchain para garantizar la confiabilidad y robustez de los datos sin incrementar significativamente la complejidad del firmware. Mediante el uso de una blockchain y contratos inteligentes, se ha logrado un sistema seguro y confiable que permite el control automatizado de dispositivos IoT, ofreciendo una plataforma escalable y trazable para aplicaciones futuras. A su vez, la plataforma Blink ofrece una interfaz visual para gestionar dispositivos conectados y flujos de datos

La implementación se realizó en una red blockchain de pruebas local, aunque el sistema ha sido diseñado para poder funcionar de manera equivalente en un entorno productivo. Para la construcción del sistema, se usó la herramienta scaffold-eth-2, la cual permite el desarrollo y despliegue de contratos inteligentes en Solidity en redes compatibles con Ethereum. Scaffold-eth-2 integra el framework Hardhat, que facilita el despliegue de los contratos, y cuenta con un frontend en Next.js que se conecta directamente al contrato inteligente. Esta interfaz es amigable y se adapta automáticamente al contrato, lo que permite que el usuario visualice el estado del contrato y acceda a sus métodos de forma intuitiva.

Además, el sistema incluye una API en el servidor de Next.js, que permite la comunicación entre los dispositivos IoT y la blockchain. El dispositivo ESP32 reporta datos a la blockchain mediante un POST HTTP a la API Rest, y en intervalos de tiempo definidos., el mismo dispositivo consulta a la misma API mediante un GET HTTP el estado objetivo de

la estufa, ya sea encendido o apagado, y activa su actuador (la estufa) en base a este parámetro. Este valor se almacena y actualiza en la blockchain según las lecturas de los sensores y los umbrales de temperatura previamente configurados en el contrato inteligente. De este modo, el contrato controla de manera automática el estado de la estufa: cuando el dispositivo IoT consulta el valor almacenado en la blockchain, actúa en consecuencia, encendiendo o apagando la estufa según lo indicado por el contrato.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo está formado por 3 estudiantes de grado, 2 docentes investigadores nóveles, un docente doctorando y un docente doctor.

Serán beneficiarios los estudiantes de carreras como la Licenciatura en Ciencias de la Computación de la Universidad Nacional de Cuyo, otras carreras de la Universidad Nacional de Cuyo en las cuales los integrantes del proyecto se desempeñan como profesores, como carreras dictadas en el Instituto Tecnológico Universitario (ITU), como carreras de otras unidades académicas.

Por lo mencionado en los párrafos anteriores, también serán beneficiarios los desarrolladores de aplicaciones y usuarios en general de plataformas de IoT, tanto los integrantes de la comunidad científica, los profesionales y los recursos humanos del medio local.

5. BIBLIOGRAFIA

[1] Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. IEEE Internet of Things Journal, 3(5), 637-646.

[2] Khan, S., Parkinson, S., & Qin, Y. (2020). Fog computing security: A review of current applications and security solutions. Journal of Cloud Computing, 9(1), 1-23.

[3] Dastjerdi, A. V., & Buyya, R. (2016). Fog computing: Helping the Internet of Things realize its potential. Computer, 49(8), 112-116.

[4] Wortmann, F., & Flüchter, K. (2015). Internet of Things: Technology and Value Added. Springer International Publishing.

[5] Seok-Won LeeIrish SinghMasoud Mohammadian. “Blockchain Technology for IoT Ap-plications”. ISBN: 978-981-33-4122-7. Springer. 2021.

[6] O. Marianetti, P. Godoy, E. Chediak, C. García Garino: “A contribution to security in IOT system. Reconfigurable Logic Device Technology and Fog Computing”. pp.6. in Short Papers of the 9th Conference on Cloud Computing Conference, Big Data & Emerging Topics, De Giusti, Armando et. al. Ed. Universidad Nacional de La Plata, Facultad de Informát-ica. La Plata, Buenos Aires, Argentina, June 22-25, 2021

[7] Osvaldo Marianetti, Pablo Godoy, Ernesto Chediak, Daniel Fontana and Carlos García Garino, “A Preliminary study of vehicular traffic distribution in Great Mendoza area. Use of IoT tools”. https://sedici.unlp.edu.ar.

[8] Use of Houshyar Honar Pajooh, Mohammad Rashid, Fakhrul Alam and Serge Demidenko, “Hyperledger Fabric Blockchain for Securing the Edge Internet of Things”. Sensors 2021, 21, 359. https:// doi.org/10.3390/s21020359.

[9] Adnan Iftekhar, Xiaohui Cui , Qi Tao and Chengliang Zheng, “Hyperledger Fabric Ac-cess Control System for Internet of Things Layer in Blockchain-Based Applications”. Entropy 2021, 23, 1054. https://doi.org/10.3390/e23081054.

· BDDM ·

Bases de Datos y Minería de Datos

Análisis y manejo inteligente de la información

W. Hasperué[1], L. Lanzarini[1], C. Estrebou[1], G. Camele[1,3], E. Rucci[1], F. Ronchetti[1], E. Lanza[2,4], R. Bértoli[2], A. J. Lira[2], A. Fernandez Bariviera[5]

[1] Instituto de Investigación en Informática LIDI*, Facultad de Informática, UNLP, La Plata, Argentina
[2] Facultad de Informática, Universidad Nacional de La Plata, La Plata, Argentina
[3] Becario postgrado UNLP.
[4] Facultad de Ingeniería, Universidad Austral, Buenos Aires, Argentina
[5] Dpto. de Economía, Universitat Rovira i Virgili, Reus, España
*** Centro asociado de la Comisión de Investigaciones Científicas de la Pcia. De Bs. As. (CIC)**

{whasperue, laural, cesarest, gcamele, erucci, fronchetti}@lidi.info.unlp.edu.ar, emlanza@mail.austral.edu.ar, {bertoli.rafa, alira}@sedici.unlp.edu.ar, aurelio.fernandez@urv.net

CONTEXTO

Esta presentación corresponde a las tareas de investigación que se llevan a cabo en el III-LIDI en el marco del proyecto "Inteligencia de Datos. Técnicas y Modelos de Machine Learning" perteneciente al Programa de Incentivos (2023-2026).

RESUMEN

Esta línea de investigación se centra en el estudio y desarrollo de Sistemas Inteligentes para la resolución de problemas de Big Data y Minería de Datos utilizando técnicas de Aprendizaje Automático. Los sistemas desarrollados se aplican particularmente al procesamiento de grandes volúmenes de información y al procesamiento de flujo de datos.

Las investigaciones correspondientes al procesamiento de datos masivos están enfocadas en el estudio y desarrollo de técnicas de selección de características, donde el foco está puesto en la reducción de los tiempos de cómputo. La optimización puede realizarse tanto en la mejora de la ejecución en un entorno distribuido, como en la propuesta de técnicas metaheurísticas que permitan obtener un subconjunto óptimo de atributos.

Por otro lado, y desde el punto de vista de la salud, se está trabajando con investigadores del CENEXA (CONICET-UNLP-CIC) en la obtención de modelos de predicción de diabetes y prediabetes.

También se investigan modelos predictivos aplicables a series temporales. En particular, el caso de estudio actual es la predicción de la latencia de respuesta en arquitecturas cloud-native utilizando modelos de machine learning. La adopción de tecnología de contenedores está creciendo rápidamente debido a su eficiencia, portabilidad y capacidad para facilitar el uso de microservicios sin embargo gestionar sistemas a gran escala es una tarea compleja. Con estos modelos se espera reducir los tiempos de respuesta de las aplicaciones.

Palabras clave: Big Data, Minería de Datos, Diabetes, Contenedores, Microservicios, Reducción de características, Tiny ML, Transformers, MapReduce, Repositorios digitales.

1. INTRODUCCION

El Instituto de Investigación en Informática LIDI tiene una larga trayectoria en el estudio, investigación y desarrollo de Sistemas Inteligentes basados en distintos tipos de estrategias adaptativas. Los resultados obtenidos han sido aplicados en la solución de problemas de distintas áreas, como salud, bioinformática, orquestación de contenedores, detección de bots y sistemas embebidos. A continuación, se detallan las investigaciones realizadas durante el último año.

1.1. BIG DATA

Selección de características

El objetivo de los algoritmos de selección de características es el de reducir las entradas a un tamaño apropiado para su procesamiento y análisis. La selección de características implica la elección de ciertos atributos, tal que, con ese subconjunto, las "propiedades naturales" de los datos no sean alteradas.

En el último año se finalizó con un doctorado, cuyo tema principal trató en distribuir de manera óptima el cómputo entre los diferentes nodos de un cluster de computadoras. Las tareas a distribuir consisten en el entrenamiento y evaluación de modelos de machine learning. Cada entrenamiento-evaluación es un agente de la metaheurística poblacional Black Hole Algorithm.

En trabajos previos se midieron los tiempos de ejecución requeridos por diferentes técnicas de machine learning [1] [3] [4]. Los resultados de este doctorado fueron enviados para publicar y aún se encuentra en evaluación.

Construcción de una herramienta de cátedra para el estudio del framework MapReduce

En esta línea se finalizó con la implementación de una herramienta para la cátedra "Conceptos y Aplicaciones de Big Data". Esta herramienta es una aplicación web que facilita a los estudiantes la creación de clústeres, la definición de Jobs MapReduce y posibilita una ejecución distribuida a través de internet utilizando las computadoras de los estudiantes logueados a la herramienta, los cuales actuarían como nodos esclavos del clúster.

Es importante aclarar que esta herramienta no busca reemplazar el framework MapReduce. Sino que, de una manera amigable, permita a los estudiantes (en el aula o en sus casas) analizar el paradigma, estudiando cómo se dividen las tareas, con cuánta información trabaja cada nodo, cuántos datos se transmiten en el clúster, etc. Resultado de este desarrollo es una tesis de grado y una publicación [8].

Detección de Bots en Repositorios Digitales

SEDICI, el repositorio digital académico de la Universidad Nacional de La Plata, es una plataforma clave para la difusión de publicaciones científicas y académicas. Sin embargo, como muchos repositorios digitales, enfrenta el desafío de distinguir entre accesos humanos y automatizados (bots), lo que afecta la precisión de sus estadísticas de uso y plantea riesgos para la seguridad cibernética. Esta problemática motivó la investigación y desarrollo de alternativas para mejorar la detección de bots mediante técnicas de aprendizaje automático.

Un aporte significativo es la creación de un nuevo dataset basado en logs web de SEDICI, que incluye información detallada sobre los accesos. Este conjunto de datos, busca subsanar la escasez de recursos al momento de evaluar distintas soluciones en este campo para facilitar futuras investigaciones.

En cuanto a los métodos, se evaluaron diversos algoritmos de aprendizaje automático, desde técnicas clásicas como la Regresión Logística hasta enfoques avanzados como XGBoost y Random Forest. Los resultados demostraron que los modelos basados en árboles y técnicas de ensemble, particularmente Random Forest y XGBoost, destacaron por su alta efectividad y capacidad para manejar el desbalance entre las clases (humanos vs. bots) [7].

Actualmente, se está implementando y probando estos algoritmos en el sitio de SEDICI, con el fin de obtener una comprensión más precisa del impacto real de las publicaciones académicas. Además, se planea ampliar el dataset existente y explorar técnicas de aprendizaje profundo, contribuyendo al estado del arte en la detección de bots y ofreciendo soluciones para repositorios digitales académicos.

Identificación de personas con riesgo de diabetes y prediabetes

La Diabetes Tipo 2 (DT2) es una enfermedad crónica que afecta a la regulación del nivel de

azúcar en sangre, siendo un serio problema de salud pública. En ese sentido, resulta importante reconocer que las consecuencias negativas de esta enfermedad comienzan en una etapa previa conocida como prediabetes, la cual implica un riesgo elevado de desarrollar DT2 en los siguientes años [6].

El desarrollo de la DT2 es un proceso lento y progresivo condicionado por factores genéticos, ambientales y de comportamiento. Aunque no existe una cura definitiva para esta enfermedad, varios estudios han demostrado que se puede prevenir o demorar su aparición. Sin embargo, la detección de DT2 y prediabetes representa un verdadero desafío para la medicina debido a la ausencia de síntomas patogenómicos y/o la falta de conocimiento de los factores de riesgo asociados.

Los modelos existentes para predicción de diabetes y prediabetes no necesariamente aplican a la población argentina [2] y, hasta donde llega nuestro conocimiento, no existe modelo ni herramienta similar disponible en nuestro medio que permita identificar personas con alta probabilidad de tener estas enfermedades. Esta línea propone desarrollar y validar modelos predictivos de diabetes y prediabetes específicos para la población argentina utilizando técnicas de Aprendizaje Automático. En particular, ya se ha iniciado el trabajo experimental contando con resultados preliminares prometedores [5][9].

1.2. PREDICCIÓN DE LATENCIA EN MICROSERVICIOS

La adopción de la tecnología de contenedores está creciendo rápidamente. Los desarrolladores de software encuentran en los contenedores una herramienta que les permite enfocarse en la aplicación en lugar de adaptar diferentes entornos de servidores. Estos son compactos, se ejecutan rápidamente, operan en cualquier entorno, pueden dividirse en módulos pequeños y son autosuficientes. En sistemas a gran escala, gestionar manualmente aplicaciones con cientos o miles de contenedores es complejo, por lo que la orquestación de contenedores es esencial.

Los contenedores facilitan el uso de microservicios, un enfoque arquitectónico que divide las aplicaciones en servicios pequeños e independientes. Estos se comunican mediante APIs bien definidas, brindando flexibilidad, seguridad y permitiendo que equipos autónomos los gestionen. Sin embargo, el rendimiento de los microservicios en un clúster puede degradarse de manera impredecible, generalmente manifestándose en mayores tiempos de respuesta, un indicador clave de confiabilidad.

Esta línea de investigación tiene por objetivo predecir la latencia de respuesta de extremo a extremo en una arquitectura de cloud-native (microservicios) sobre un caso de estudio concreto. Las tareas están siendo realizadas en el marco de una tesis correspondiente a la carrera de Maestría en Ciencia de Datos de la Universidad Austral (CABA). Actualmente se está trabajando en el desarrollo de un modelo que sea capaz de predecir la latencia futura de un microservicio usando múltiples variables, estudiando especialmente los modelos basados en Transformers que, si bien han demostrado ser útiles para anticipar comportamientos inestables o para la comprensión de texto, aún no se han utilizado para predecir latencias. El énfasis está puesto en comparar estas arquitecturas con métodos clásicos ampliamente utilizados para esta predicción, utilizando una implementación de referencia de microservicios para el benchmarking de aplicaciones. Los resultados obtenidos hasta el momento han sido publicados en [14].

1.3. TINYML

Hace varios años se viene desarrollando un proyecto de investigación centrado en TinyML [10, 11], con el objetivo de adaptar técnicas de aprendizaje automático a dispositivos de hardware limitado, como microcontroladores (MCUs). Este enfoque busca implementar inteligencia artificial en dispositivos con restricciones de memoria, capacidad de procesamiento y consumo energético.

El núcleo del proyecto es EmbedIA, un framework de código abierto que transforma modelos de aprendizaje automático y redes neuronales en código C/C++/Arduino, permitiendo su ejecución en cualquier MCU

sin requerir hardware especializado. A diferencia de otras soluciones que dependen de dispositivos de 32 bits o soporte para instrucciones DSP/SIMD, EmbedIA es compatible con una amplia variedad de microcontroladores. Además, ha demostrado ser más eficiente que otros frameworks en términos de uso de memoria y velocidad de ejecución [11, 12, 13].

EmbedIA posee varias funcionalidades como Tensorflow y TFlite, redes binarias Larq y modelos Scikit-Learn basados en máquinas de soporte vectorial (SVM), árboles de decisión, el algoritmo KNN, y se está avanzando en la integración de Random Forest.

Actualmente, se está ampliando la compatibilidad de EmbedIA con formatos como PyTorch y ONNX, para permitir una mayor flexibilidad en la integración de modelos más allá de Tensorflow y TFLite. Además, se están probando y adaptando modelos optimizados para tareas específicas, como el reconocimiento de objetos, en placas de desarrollo con MCUs. Paralelamente, se están incorporando soporte para más capas Tensorflow y algoritmos de aprendizaje automático. También se tiene planificada la exploración de MCUs con unidades de procesamiento neuronal (NPUs), una tecnología emergente que promete mejorar el rendimiento en tareas de inferencia. Estas pruebas buscan evaluar cómo EmbedIA puede aprovechar estas capacidades para optimizar aún más el procesamiento en dispositivos con recursos limitados.

2. TEMAS DE INVESTIGACIÓN Y DESARROLLO

- Medición de performance de algoritmos de machine learning en entornos distribuidos.
- Módulo de balance de carga de tareas entre nodos en un entorno Spark.
- Desarrollo de una herramienta de cátedra que permite estudiar el funcionamiento del framework MapReduce.
- Identificación de Personas con Riesgo de Diabetes y Prediabetes
- Estudio de modelos de Deep Learning para series temporales.
- Estudio de aplicaciones y bibliotecas diseñadas para realizar benchmarks de modelos de deep learning aplicados a series temporales.
- Estudio de técnicas de compresión de modelos para microcontroladores.
- Análisis de bibliotecas y frameworks de aprendizaje automático para microcontroladores.
- Desarrollo de un Framework de código abierto para ejecución de modelos de aprendizaje automático y redes neuronales en microcontroladores.
- Exploración de microcontroladores con unidades de procesamiento neuronal (NPUs).

3. RESULTADOS OBTENIDOS

- Medición de tiempos de ejecución de algoritmos de ML en un entorno Spark.
- Un emulador de MapReduce capaz de funcionar a través de internet.
- Modelos de predicción de diabetes y prediabetes
- Análisis comparativo de modelos de Deep Learning para series temporales.
- Desarrollo de un modelo de predicción de latencia en una aplicación basada en microservicios.

4. FORMACIÓN DE RECURSOS HUMANOS

El grupo de trabajo de la línea de I/D aquí presentada está formado por: cuatro profesores doctores con dedicación exclusiva, un profesor con dedicación exclusiva, un alumno diplomado, tres tesistas de Doctorado en Cs. Informáticas (uno con beca de postgrado de la UNLP) y un profesor extranjero.

En el último año se han finalizado 2 tesis de doctorado, 2 tesis de especialista, 2 tesinas de grado de Licenciatura y 2 prácticas profesionales supervisadas.

Actualmente se están desarrollando 3 tesis de doctorado, 5 tesis de maestría, 3 tesinas de Licenciatura y 2 prácticas profesionales supervisadas. También participan en el desarrollo de las tareas becarios y pasantes del III-LIDI.

5. REFERENCIAS

[1] Camele, G.; Hasperué, W. Performance analysis of the Survival-SVM classifier applied to gene expression databases. XXIX Congreso Argentino de Ciencias de la Computación. Univ. Nac. de Luján. 2023.

[2] Choudhury A., Gupta D. A Survey on Medical Diagnosis of Diabetes Using Machine Learning Techniques. In: Recent Developments in Machine Learning and Data Analytics. Advances in Intelligent Systems and Computing, vol 740. Springer, Singapore. 2019.

[3] Camele, G.; Hasperué, W.; Ronchetti, F.; Quiroga, F.; A comparative study of the performance of four classification algorithms from the Apache Spark ML library. CACIC. Salta. 2021.

[4] Camele, G.; Hasperué, W.; Ronchetti, F.; Quiroga, F.M. Statistical analysis of the performance of four Apache Spark ML algorithms. Journal of Computer Science and Technology 22 (2), e14-e14, 2022.

[5] Enzo Rucci, Gonzalo Tittarelli, Franco Ronchetti, Jorge F. Elgart, Laura Lanzarini, Juan José Gagliardino. "Primeras Experiencias en la Identificación de Personas con Riesgo de Diabetes en la Población Argentina utilizando Técnicas de Aprendizaje Automático". XXIX Congreso Argentino de Ciencias de la Computación. Octubre de 2023.

[6] Federación Internacional de Diabetes, "Atlas de la diabetes de la fid," tech. rep., Federación Internacional de Diabetes, 2017. https://diabetesatlas.org/resources/2017-atlas.html.

[7] Bértoli R., Estrebou C., Lira A. J. Aprendizaje Automático para la Detección de Bots en Repositorios Digitales. XXX Congreso Argentino de Ciencias de la Computación – CACIC 2024. Págs. 12-23. ISBN: 978-950-34-2428-5. Octubre 2024.

[8] Scoffield, D., Giorgetti, V., Hasperué, W. DOMEX: un emulador del framework Map-Reduce. XXX Congreso Argentino de Ciencias de la Computación. Octubre de 2024.

[9] Rucci, E., Tittarelli, G., Ronchetti, F., Elgart, J. F., Lanzarini, L. y Gagliardino, J. J. First Experiences with the Identification of People at Risk for Diabetes in Argentina Using Machine Learning Techniques, en Computer Science – CACIC 2023, P. Pesado, W. Panessi, y J. M. Fernández, Eds., Cham: Springer Nature Switzerland, 2024, pp. 227-239. doi: 10.1007/978-3-031-62245-8_16

[10] Estrebou C., Feming M., Saavedra M. D., Adra F. MbedML: A Machine Learning Project for Embedded Systems. IX Jornadas de Cloud Computing, Big Data & Emerging Topics. ISBN: 978-950-34-2016-4 Pp. 25-28. Facultad de Informática. UNLP. Junio 2021.

[11] Estrebou, C., Saavedra, M. D., Adra, F., Fleming, M. TinyML for Small Microcontrollers. X Jornadas de Cloud Computing, Big Data & Emerging Topics. ISBN 978-950-34-2126-0. Pp 42-46. Facultad de Informática. UNLP. Mayo 2022.

[12] Estrebou C., Feming M., Saavedra M. D., Adra F. A Neural Network Framework for Small Microcontrollers. XXVII Congreso Argentino de Ciencias de la Computación. ISBN: 978 -987-633-574-4. Pp. 51-60. Univ. Nacional de Salta. Octubre 2021.

[13] Estrebou C., Feming M., Saavedra M. D., Adra F., De Giusti, A. E. Lightweight Convolutional Neural Networks Framework for Really Small TinyML Devices. Second International Conference on Smart Technologies, Systems and Applications. SmartTech-IC 2021. Quito, Ecuador. Diciembre 2021.

[14] Lanza, E., Lanzarini L., Estrebou, C. Predicción de Latencia en Microservicios con Modelos de Deep Learning. XXX Congreso Argentino de Ciencias de la Computación – CACIC 2024. Octubre de 2024.

Aplicación de Algoritmos de detección de Outliers en el Área de Salud

María Fernanda Cuadrado Estrebou, Alejandro Luis Foglino, María Teresa Peña

Universidad Argentina de la Empresa (UADE)
{mcuadrado, afoglino, mapena}@uade.edu.ar

RESUMEN

Esta investigación aborda el desafío de procesar grandes volúmenes de información en áreas como la salud, donde la gestión eficiente de los datos es crucial para la toma rápida de decisiones. Uno de los desafíos más comunes es garantizar la confiabilidad de la información disponible, evitando sesgos introducidos por la presencia de datos atípicos que puedan influir negativamente en las decisiones. En este contexto, la detección de *outliers* [1] adquiere especial relevancia, entendiéndose como tales aquellos valores atípicos que pueden surgir debido a errores en los equipos de medición, registros incorrectos u otros factores que afectaron la captura o almacenamiento de los datos.

En particular, esta investigación se centra en el análisis de datos provenientes de estudios clínicos realizados en un laboratorio hospitalario. Se aplicarán distintos algoritmos de detección de *outliers* con el objetivo de identificar cuál de ellos se ajusta mejor a este tipo de datos y qué consideraciones o ajustes deben realizarse para su correcta implementación.

Palabras Clave: outliers, confiabilidad, grandes bases de datos, eficiencia, toma de decisiones.

CONTEXTO

El análisis de datos en dominios complejos que manejan grandes volúmenes de información representa un desafío significativo para los profesionales que trabajan en estos entornos, y el sector de la salud no es la excepción.

El procesamiento de datos en laboratorios de análisis clínicos es un proceso fundamental para garantizar la calidad y precisión en los diagnósticos médicos. Sin embargo, la presencia de valores atípicos (*outliers*) en los datos puede comprometer la interpretación de los resultados, llevando a decisiones erróneas o inexactas. Estos valores atípicos pueden originarse por errores en la medición, fallas en los equipos, errores humanos en la transcripción de datos o incluso condiciones clínicas raras que requieren una evaluación especial. Identificar y manejar adecuadamente estos valores es crucial para mejorar la confiabilidad de los análisis y la toma de decisiones médicas basadas en ellos.

Dado este desafío, diversas metodologías han sido desarrolladas para la detección de *outliers* en conjuntos de datos clínicos. Entre ellas, los enfoques estadísticos tradicionales, como el uso de desviaciones estándar y pruebas de hipótesis, han sido ampliamente aplicados los siguientes métodos:

- El método de puntuación Z[6] es una técnica estadística comúnmente utilizada para identificar valores atípicos. Se basa en calcular cuántas desviaciones estándar separan un valor de la media de la distribución.
- El rango intercuartílico[6] es otra técnica estadística empleada para detectar *outliers*, método que se basa en calcular la diferencia entre el tercer cuartil y el primer cuartil (Q1) de los datos.

Estudios previos han abordado la aplicación de estas metodologías en diferentes dominios de la ciencia de datos. Por ejemplo, investigaciones han demostrado la eficacia de los métodos basados en árboles de decisión para la detección

de anomalías en registros electrónicos de salud, así como la utilidad de enfoques basados en clustering en la identificación de patrones inusuales en estudios epidemiológicos. Además, trabajos recientes han explorado el uso de redes neuronales para la identificación de valores atípicos en datos clínicos, con resultados prometedores en términos de precisión y capacidad de generalización.

Sin embargo, con el crecimiento en la cantidad y complejidad de los datos generados en los laboratorios, se han desarrollado técnicas más avanzadas basadas en aprendizaje automático y minería de datos. Algoritmos como Isolation Forest, Local Outlier Factor (LOF) y One-Class SVM [7] han mostrado un alto rendimiento en la identificación de valores atípicos en grandes volúmenes de datos, permitiendo una mayor precisión y adaptabilidad a diferentes tipos de distribuciones.

A pesar de estos avances, la aplicación de algoritmos de detección de *outliers* en el contexto específico de laboratorios de análisis clínicos sigue siendo un área de investigación en evolución. La heterogeneidad de los datos clínicos, la variabilidad entre distintos laboratorios y la necesidad de minimizar falsos positivos son desafíos que requieren un estudio detallado. Este trabajo busca contribuir a esta línea de investigación evaluando y comparando distintos algoritmos de detección de *outliers* aplicados a datos reales de un laboratorio de análisis clínico, con el objetivo de determinar cuáles son más efectivos y robustos para este tipo de entorno.

1. INTRODUCCIÓN

En la actualidad, los laboratorios de análisis clínicos generan una gran cantidad de datos que son utilizados para diagnosticar enfermedades, monitorear tratamientos y realizar estudios epidemiológicos. La precisión y confiabilidad de estos datos son fundamentales para la toma de decisiones en el ámbito médico, ya que cualquier desviación o error en los resultados puede afectar directamente la salud de los pacientes. Sin embargo, dentro de estos conjuntos de datos es común encontrar valores atípicos (*outliers*), los cuales, como se ha mencionado anteriormente, pueden ser el resultado de errores en la medición, fallas en los equipos, errores humanos o incluso condiciones clínicas poco comunes. La detección y el tratamiento adecuado de estos valores atípicos es esencial para mejorar la calidad del análisis y reducir el riesgo de interpretaciones erróneas.

La identificación de valores atípicos ha sido un problema ampliamente estudiado en diversas disciplinas, dando lugar a diferentes enfoques metodológicos. Los métodos estadísticos tradicionales, como la detección basada en la desviación estándar o el rango intercuartílico, han sido ampliamente utilizados en el análisis de datos clínicos. No obstante, el crecimiento exponencial de la cantidad de datos generados en los laboratorios clínicos, junto con la necesidad de una detección más precisa y adaptable, ha impulsado el desarrollo y aplicación de técnicas avanzadas de aprendizaje automático. Algoritmos como Isolation Forest, Local Outlier Factor (LOF) y One-Class SVM[7] han demostrado ser altamente efectivos para la detección de anomalías en grandes volúmenes de datos, proporcionando un enfoque más robusto en comparación con los métodos convencionales. Cada uno de los algoritmos mencionados se caracteriza por el siguiente comportamiento principal:

- Isolation Forest es un algoritmo basado en árboles de decisión diseñado específicamente para detectar anomalías. Su enfoque tiene como principio que los outliers son más fáciles de aislar que los datos normales.
- LOF en cambio es un algoritmo basado en densidad que compara la densidad local de un punto con la de sus vecinos para determinar el outlier.
- One-Class SVM es una variante del algoritmo SVM utilizada para la detección de anomalías, aprendiendo un modelo de frontera que separa los datos normales de posibles outliers.

A pesar de los avances en la detección de *outliers*, la aplicación de estos algoritmos en el ámbito de los laboratorios de análisis clínicos

sigue siendo un área en desarrollo. La variabilidad de los datos clínicos, la heterogeneidad de los métodos de análisis y la necesidad de minimizar los falsos positivos son desafíos que requieren estudios más detallados para garantizar la eficacia de estas metodologías en entornos reales. En este contexto, el presente trabajo tiene como objetivo evaluar y comparar distintos algoritmos de detección de *outliers* aplicados a datos reales de un laboratorio de análisis clínico. Se analizará la efectividad de estos métodos en la identificación de valores atípicos, considerando métricas de precisión, sensibilidad y especificidad, con el fin de determinar cuáles ofrecen un mejor desempeño en este tipo de datos.

Este estudio no solo busca aportar un análisis comparativo entre distintas metodologías, sino también generar lineamientos para su implementación en entornos clínicos reales. Una detección precisa de valores atípicos en los datos clínicos no solo mejora la calidad de los resultados, sino que también contribuye a incrementar la seguridad del paciente y la eficiencia en la toma de decisiones médicas. A través de esta investigación, se espera proporcionar herramientas que permitan a los laboratorios clínicos adoptar estrategias más eficaces para la detección de anomalías, facilitando así un procesamiento de datos más preciso y confiable.

2. LINEAS DE INVESTIGACIÓN Y DESARROLLO

Este proyecto se enmarca dentro de una línea de investigación dedicada al tratamiento de grandes bases de datos y la aplicación de herramientas de inteligencia artificial para la optimización de procesos analíticos y la mejora en la toma de decisiones. En continuidad con el trabajo *Algoritmos de Detección de Outliers* [4], donde se exploraron estrategias para optimizar distintos algoritmos con el fin de mejorar su desempeño computacional, el presente estudio amplía el enfoque hacia la evaluación integral de su efectividad y eficiencia. Específicamente, se busca no solo analizar el impacto de las optimizaciones en términos de tiempos de procesamiento y consumo de recursos, sino también evaluar su capacidad para detectar anomalías con precisión en distintos contextos y conjuntos de datos. Esta aproximación permite una comprensión más profunda de los beneficios y limitaciones de cada método, facilitando su aplicación en entornos reales y su adaptación a diferentes necesidades analíticas.

3. RESULTADOS OBTENIDOS/ESPERADOS

El objetivo de esta investigación es identificar los algoritmos de detección de *outliers* más adecuados para el dominio de estudio, en este caso, laboratorios de análisis clínicos. Dado que la precisión en la identificación de valores atípicos es crucial en este ámbito, se busca no sólo determinar qué métodos ofrecen el mejor desempeño, sino también establecer criterios específicos para su implementación eficiente. Esto implica analizar la sensibilidad y especificidad de cada algoritmo en distintos tipos de pruebas de laboratorio, considerando variaciones en la calidad y volumen de los datos procesados.

Además, se pretende evaluar la forma óptima de aplicar estos algoritmos, identificando sus ventajas y desventajas en función de factores como la complejidad computacional, la interpretabilidad de los resultados y la facilidad de integración en los sistemas utilizados por los laboratorios clínicos. Para garantizar una evaluación rigurosa, se trabaja en colaboración con un grupo de profesionales de la salud, quienes, a partir de su experiencia en el área, analizarán la efectividad de los algoritmos y su aplicabilidad en escenarios reales. Su retroalimentación permitirá no solo validar la utilidad práctica de cada enfoque, sino también identificar posibles limitaciones y áreas de mejora en su implementación.

Por otro lado, los resultados obtenidos se comparan con métodos estadísticos tradicionales, como el cálculo de intervalos de confianza, a fin de determinar su validez y aplicabilidad en este contexto específico. Esta comparación permitirá establecer si los enfoques clásicos siguen siendo herramientas confiables para la detección de *outliers* en análisis clínicos

o si es necesario recurrir a técnicas más avanzadas para mejorar la precisión y confiabilidad de los diagnósticos.

4. FORMACIÓN DE RECURSOS HUMANOS

Los participantes de esta investigación incluyen docentes de informática y matemáticas/estadística, así como estudiantes de grado en informática. Los estudiantes recibirán formación en la implementación de algoritmos, con un enfoque en el lenguaje Python, además de adquirir conocimientos fundamentales en estadística y matemáticas. Paralelamente, los docentes de ambas disciplinas compartirán sus conocimientos especializados, promoviendo un aprendizaje multidisciplinario y enriqueciendo la colaboración entre ambas áreas.

5. BIBLIOGRAFÍA

[1] Detection in Large Database. 2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery [en línea]. Tianjin, China: IEEE, pp. 495-499. [Consulta: 9 junio 2022]. ISBN 978-0-7695-3735-1. DOI 10.1109/FSKD.2009.303.

[2] BRADLEY, Andrew P., 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognition, vol. 30, no. 7, pp. 1145-1159. ISSN 0031-3203. DOI 10.1016/S0031-3203(96)00142-2.

[3] HAWKINS D. Identification of Outliers (Monographs on Statistics and Applied Probability) vol 3. Chapman and Hall, London. 1980.

[4] CUADRADO, De ARMAS, FOGLINO, Algoritmos de detección de Outliers, WICC 2022.

[5] KISHAN G. MEHROTA C., HUAMING. Anomaly Detection Principles and Algorithms. 2017

[6] MANAI, SABRI. Detección de anomalías en series temporales univariantes. Trabajo Final de Master-2024.

[7] SADEQ, DARRAB y Otros. Anomaly Detection Algorithms: Comparative Analysis and Explainability Perspectives. 2023.

Aplicaciones de Bases de Datos Espaciales y Espacio Temporales

Javier Apolloni, María Gisela Dorzán, Pablo Palmero y María Teresa Taranilla

Departamento de Informática
Facultad de Ciencias Físico Matemáticas y Naturales
Universidad Nacional de San Luis, Argentina
{javierma, mgdorzan, prpalmero, tarani}@unsl.edu.ar

RESUMEN

En el *Proyecto Tecnologías Avanzadas de Bases de Datos* de la Universidad Nacional de San Luis se enmarca la línea de investigación *Bases de Datos Espaciales y Espacio Temporales*. Esta línea de investigación se enfoca en el estudio de técnicas y herramientas tanto para la investigación de base, así como en el desarrollo de aplicaciones abordando problemáticas en diversas áreas de aplicación de las bases de datos espaciales y espacio temporales.

Palabras clave: Bases de Datos, Bases de Datos Espaciales y Espacio Temporales, Geometría Computacional, Trazabilidad, Gestión Integral de Incendios Forestales

CONTEXTO

La línea de investigación *Bases de Datos Espaciales y Espacio Temporales* forma parte del Proyecto *Tecnologías Avanzadas de Bases de Datos* (PROICO 03-1723) integrado en el Programa de Incentivos y en el contexto del *Laboratorio de Investigación y Desarrollo en Base de Datos (LaBDa)* de la Facultad de Ciencias Físico Matemáticas y Naturales de la Universidad Nacional de San Luis.

El proyecto se compone de tres líneas de investigación centradas en el estudio de modelos de bases de datos no convencionales, y en el desarrollo de estrategias para gestionar y recuperar información almacenada en repositorios de datos no estructurados.

Con el avance de las tecnologías de información y comunicación surge la necesidad de utilizar modelos de bases de datos con capacidad para manejar información compleja en escenarios espaciales y espacio temporales. En algunos entornos de trabajo, las bases de datos tradicionales resultan insuficientes para representar fenómenos que varían en el tiempo y el espacio, y son necesarias nuevas estrategias para el almacenamiento, consulta y análisis de datos.

En este contexto, la línea de investigación en Bases de Datos Espaciales y Espacio Temporales se propone estudiar técnicas y herramientas que faciliten la gestión y extracción de información. Esto abarca el diseño de aplicaciones para modelar y administrar datos, así como la obtención de información para realizar análisis complejos y generar nuevo conocimiento que respalde la toma de decisiones en diferentes sistemas. El trabajo de investigación se lleva a cabo en colaboración con investigadores de otros proyectos en la Universidad Nacional de San Luis.

1. INTRODUCCIÓN

Las Tecnologías de la Información y la Comunicación (TIC) han tenido un avance significativo, transformando la manera en que las personas trabajan, interactúan y se comunican a nivel global. En diversos ámbitos,

se observan experiencias que promueven prácticas orientadas a optimizar tareas relacionadas con la gestión de datos en diferentes temáticas. Asimismo, tecnologías tales como los Sistemas de Posicionamiento Global (GPS), los dispositivos móviles y los sensores remotos, entre otras, han generado grandes volúmenes de datos que registran la ubicación de objetos, sus movimientos y su evolución en el tiempo. Este enorme caudal de datos presenta desafíos respecto de su almacenamiento, consulta, análisis y visualización.

En diversas aplicaciones, la generación de información confiable es esencial para la toma de decisiones, especialmente cuando se manipulan datos espaciales y temporales, en los que la ubicación geográfica y el tiempo son claves para comprender fenómenos complejos y dinámicos. Las bases de datos espaciales permiten almacenar la información de la ubicación espacial de los objetos, sus características y las relaciones espaciales entre ellos. Las bases de datos temporales agregan el componente temporal, lo que permite organizar la información y almacenar tanto datos actuales como datos históricos. Las bases de datos espacio temporales combinan ambas características, permitiendo analizar la evolución de objetos y eventos en el espacio a lo largo del tiempo, un aspecto fundamental para el análisis de fenómenos como el seguimiento de tendencias o el monitoreo de actividades [3] [9] [12].

La integración de las TIC con las bases de datos espaciales y espacio temporales constituye una herramienta clave en la gestión y análisis de datos espaciales y temporales. Estas tecnologías permiten almacenar, gestionar y analizar información geográfica y temporalmente referenciada, facilitando la toma de decisiones en una amplia variedad de aplicaciones, tales como la planificación urbana, la gestión de recursos naturales, la logística, agricultura y ganadería, seguridad, entre otras.

La producción agropecuaria representa un área donde la tecnología ha logrado un impacto significativo. En el marco de un convenio entre el Instituto Nacional de Tecnología Agropecuaria (INTA-Estación Experimental Agropecuaria San Luis) y la Universidad Nacional de San Luis, se creó *Campo Conectado*, un espacio dedicado a la investigación, intercambio y desarrollo de aplicaciones, con el objetivo de abordar problemáticas del sector y generar contribuciones en este ámbito. En este ámbito se identificaron diversos requerimientos relacionados con la producción ganadera, para los cuales el grupo de investigación ha propuesto algunas soluciones [1].

En el contexto del proyecto de investigación *Gestión Integral de Incendios Forestales* se aborda el estudio de incendios forestales, que representan una grave amenaza con consecuencias devastadoras, incluyendo la pérdida de vidas, biodiversidad y suelos, así como daños económicos y ambientales. En este proyecto se propone un enfoque multidisciplinario para estudiar la problemática asociada a la gestión de incendios, combinando investigación básica y aplicada, con el objetivo de desarrollar herramientas que permitan mejorar la toma de decisiones y el análisis del impacto de los incendios, facilitando una gestión eficiente y precisa. En esta iniciativa, los miembros de la línea de investigación participan activamente [11].

Por otro lado, en el ámbito de la seguridad y la prevención del delito, disponer de información precisa sobre la ubicación y los desplazamientos de vehículos en una zona específica resulta fundamental para el análisis de eventos como robos, desapariciones de personas y tráfico ilegal de bienes. En este sentido, se estudia la implementación de un sistema de información que permita obtener la trazabilidad de los vehículos, mediante el uso de los recursos tecnológicos disponibles en la provincia de San Luis.

En los ámbitos mencionados previamente, surge la necesidad de representar y gestionar tanto datos tradicionales como datos complejos.

Esto implica abordar problemáticas relacionadas con el almacenamiento y consulta de información actual e histórica, así como la representación de objetos de estudio

referenciados espacialmente, junto con su evolución temporal. Estas áreas de trabajo abordan una amplia gama de desafíos en los que las bases de datos espaciales y espacio temporales pueden realizar aportes.

2. LÍNEA DE INVESTIGACIÓN

En la línea de investigación se exploran diversos dominios de aplicación de bases de datos espaciales y espacio temporales, y se utilizan técnicas de indexación y herramientas especializadas para resolver problemas específicos.

Entre los temas de estudio, se destaca el diseño y la implementación de índices espaciales y espacio temporales adaptados en diferentes escenarios. Se trabaja en el desarrollo de herramientas y aplicaciones que optimicen el rendimiento y la utilidad de las bases de datos espaciales y espacio temporales en distintos ámbitos. En este contexto, se incorporan conceptos y técnicas de la disciplina Geometría Computacional para abordar aspectos particulares de los problemas analizados, aportando soluciones innovadoras y eficientes [2].

Los objetivos específicos propuestos en la línea son los siguientes:

- Estudiar la indexación espacial y espacio temporal aplicada a objetos en movimiento en distintos escenarios. Implementar estructuras de almacenamiento y algoritmos de consulta y de evaluación experimental.
- Aplicar bases de datos espaciales y espacio temporales en la resolución de problemas reales.
- Diseñar e implementar aplicaciones y herramientas informáticas que permitan a los usuarios acceder a soluciones prácticas accesibles en diversas plataformas y dispositivos.

3. RESULTADOS OBTENIDOS

En el proyecto *Campo Conectado* se destacó la importancia de desarrollar herramientas tecnológicas para respaldar la gestión eficiente de la producción agropecuaria. En particular, en el ámbito de la ganadería, se identificaron varios requerimientos como el seguimiento de indicadores productivos y ambientales, la trazabilidad animal y el control de la alimentación del ganado. Para abordar estos requerimientos se propuso la implementación de una plataforma integral para soportar eventos y sistemas de información vinculados con la producción ganadera.

Esta plataforma fue diseñada para realizar un seguimiento espacial y temporal de rodeos ganaderos, con el objetivo de facilitar la gestión tanto tradicional como espacial de los objetos de un establecimiento agropecuario [4] [6]. Para ello, se desarrolló una herramienta que permite el seguimiento individual de los animales, procesando datos tradicionales y espaciales. La herramienta permite resolver consultas espaciales y espacio temporales sobre los objetos de estudio.

Además, la plataforma ofrece una interfaz de comunicación con las bases de datos y la posibilidad de visualizar el seguimiento espacial de los objetos de estudio a través de dispositivos móviles. Esto facilita la toma de decisiones en tiempo real y contribuye a la gestión eficiente de los recursos disponibles.

La herramienta también se utilizó con lotes de datos del monitoreo de animales salvajes, adaptándose para visualizar y procesar información relevante sobre distintas especies, sus movimientos y comportamientos. Se realizaron consultas espacio temporales para mostrar y analizar la evolución de sus desplazamientos a lo largo del tiempo. Este enfoque amplió las capacidades de la plataforma, facilitando su adaptación a diversos contextos y necesidades de monitoreo. Este trabajo corresponde a una tesis de Maestría en Ciencias de la Computación [5] [10].

En el contexto de la seguridad, en lo que refiere a la trazabilidad de vehículos en situaciones sospechosas o ilegales, se está trabajando en la implementación de un sistema destinado a analizar la circulación de vehículos para detectar aquellos que puedan estar involucrados en actividades ilícitas. Este sistema se basa en

el seguimiento de los movimientos de los vehículos mediante la lectura de sus patentes, utilizando los recursos disponibles en la provincia de San Luis, tanto en áreas urbanas y rurales como en las fronteras provinciales. Actualmente, se continúa con el desarrollo de un prototipo para simular casos de seguimiento y con la evaluación de los recursos técnicos necesarios para su implementación. Este proyecto es un trabajo final de Licenciatura en Ciencias de la Computación [8].

Los incendios forestales amenazan vidas, propiedades y el equilibrio de los recursos naturales, afectando la economía y el medio ambiente. El proyecto *Gestión Integral de Incendios Forestales* tiene como objetivo realizar aportes que contribuyan a disminuir el riesgo y el impacto de los incendios forestales en la provincia de San Luis. En este marco, el equipo contribuye al diseño y asistencia técnica para el desarrollo de bases de datos espaciales, así como al procesamiento y optimización de consultas en las bases de datos existentes vinculadas a la problemática.

Uno de los objetivos del proyecto es el desarrollo de bases de datos para documentar y analizar la dinámica de los incendios forestales en función del tiempo y el espacio, explorando cómo la integración de datos espaciales y temporales puede ofrecer una comprensión más precisa y detallada. Se ha realizado la recopilación, análisis y procesamiento de bases de datos espaciales utilizando imágenes multiespectrales (Landsat 7-8 y 9, Sentinel 2 MSI, MODIS) y productos procesados (Landsat BA, MODIS BURN, FireCCI) relacionados con incendios forestales ocurridos en la provincia de San Luis entre 2000 y 2023. Se ha creado una base de datos en un Sistema de Información Geográfica (SIG) de software libre para gestionar esta información [12].

El grupo continúa su capacitación y estudio en teledetección y herramientas avanzadas para mejorar la gestión de incendios forestales. Además, participa en talleres y cursos especializados que abordan el uso de tecnologías para detección, monitoreo y evaluación de incendios.

Este proyecto multidisciplinario se lleva a cabo en colaboración entre la Facultad de Ciencias Físicas, Matemáticas y Naturales (FCFMyN), la Facultad de Química, Bioquímica y Farmacia (FQBF) y la Facultad de Ingeniería y Ciencias Agropecuarias (FICA) de la Universidad Nacional de San Luis. Este trabajo conjunto permite integrar diversas áreas del conocimiento, lo que enriquece el enfoque y los resultados del proyecto.

Por otra parte, se llevó a cabo el trabajo final de Licenciatura en Ciencias de la Computación titulado "*Migración del Sistema Informático Universitario Guaraní Versión 2 a la Versión 3 en la UNSL"* que se centró en la migración del sistema de gestión académica de la Universidad Nacional de San Luis, específicamente la actualización del sistema SIU guaraní de la versión 2 a la versión 3. El trabajo abarcó la migración de los sistemas de las ocho facultades de la UNSL y del Instituto Politécnico y Artístico Universitario “Mauricio A. López” (IPAU) a una única base de datos centralizada. La migración implicó actualizar la estructura de la base de datos, garantizar la compatibilidad entre ambas versiones, realizar pruebas de compatibilidad, asegurar y preservar la integridad de la información y la consistencia de los datos y la personalización de reportes para adaptarlos a los estándares de la universidad [8].

Como ejes de trabajo futuro, se plantea continuar promoviendo actividades que fomenten la relevancia y la transferencia de proyectos de investigación, desarrollo e innovación (I+D+i) en las áreas específicas de intervención.

Se espera potenciar la colaboración entre distintos equipos de trabajo, consolidar la naturaleza interdisciplinaria de la investigación y mejorar las competencias de los investigadores participantes en la línea de investigación.

4. FORMACIÓN DE RECURSOS HUMANOS

Los integrantes de la línea de investigación fortalecen su formación mediante actividades

de colaboración e intercambio de información científica y tecnológica con investigadores de la Universidad Nacional de San Luis y de otras instituciones académicas.

La formación de recursos humanos se refleja en la dirección de tesis de maestría y trabajos finales de Licenciatura en Ciencias de la Computación, así como en la dirección de becas y dictado de cursos de posgrado.

Además, se realizan actividades de formación y capacitación académica, cursos de especialización, participación en eventos de divulgación científica, conferencias y otras actividades relacionadas a la investigación.

**Este trabajo está parcialmente subvencionado por Universidad Nacional de San Luis (Argentina), PROICO 03-1723 y PICT-2021-CAT-II-00105*

5. BIBLIOGRAFÍA

[1] *Campo Conectado*, proyecto interinstitucional http://inta.gob.ar/noticias/campo-conectado-un-nuevo-proyecto-interinstitucional (2017)

[2] de Berg, M., Cheong O., van Kreveld, M., Overmars, M.: *Computational Geometry: Algorithms and Applications.*, Springer-Verlag, Heidelberg (2008)

[3] Elmasri, R. Navathe, S.B.*: Fundamentals of Database Systems.* 7ma Edition. Pearson (2015)

[4] Gagliardi, E., Dorzán, M. G., Taranilla, M. T., Palmero, P., Casanova, C.: *Propuesta de plataforma para la integración de TIC orientadas al Agro* Congreso de AgroInformática CAI 2017, 46 JAIIO (2017)

[5] Gagliardi, E.; Dorzán, M.G.; Taranilla, M.T.; Palmero, P.; Casanova, C.: *GeoSeguimiento de Rodeos, hacia una plataforma integral para el Agro*. Congreso de AgroInformática CAI 2018, 47 JAIIO (2018).

[6] Gagliardi, E..; Dorzán, M.G.; Taranilla, M.T.; Palmero, P.; Casanova, C.: *Diseño de una herramienta de apoyo en la gestión de rodeos de cría.* Congreso de AgroInformática CAI 2018, 47 JAIIO (2018).

[7] Leiva Olguin G., Martinez R., Paez B., Gagliardi E. O. Diseño de un sistema de trazabilidad de vehículos en estado de sospecha. Congreso Nacional de Ingeniería Informática / Sistemas de Información-Universidad Tecnológica Nacional (2022)

[8] Lupi Casale, Gianni; Gagliardi, E..; Dorzán, M.G; *Migración del Sistema Informático Universitario Guaraní Versión 2 a la Versión 3 en la UNSL* – Informe trabajo final de Licenciatura en Ciencias de la Computación UNSL (2024)

[9] Manolopoulos, Y., Papadopoulos, A., Vassilakopoulos, M.: *Spatial Databases: Technologies, Techniques and Trends*. Idea Group (2005)

[10] Palmero, P.; Gagliardi, E; Dorzán, M. G. *Seguimiento de rodeos en establecimientos agropecuarios.* XXVI Congreso Argentino de Ciencias de la Computación CACIC (2020)

[11] *Gestión Integral de Incendios Forestales* Proyecto de investigación de la Agencia I+D+i PICT-Aplicados Salto Institucional, PICT-2021-CAT-II-00105

[12] Shekhar, S.; Chawla, S.: *Spatial databases: a tour*. Prentice Hall (2003)

[13] *Incendios Forestales 2000-2023 en San Luis* - API en Google Earth Engine (2024)
https://lucasneoclasico.users.earthengine.app/view/areas-quemadas-firecci-sl

De los Datos al Conocimiento: Construcción de Soluciones de Analítica e Inteligencia Empresarial

María Inés Lund[1], Silvina I. Migani[1], Silvina Balmaceda[1], Cintia Ferrarini O.[1], Juan Capdevila[1**], German Alvarez[1**], Leonardo de los Rios[1**], Daniel Tapella[2], Leonardo Blasco[2], Ignacio Colomé[1*], Ramiro Savoie[3], Fernando Más[2], Cristina Vera[1], Leonardo Drazic[1*], Erica Minet Bravo[4],

[1]Departamento e Instituto de Informática, FCEFN, Universidad Nacional de San Juan
[1*]Alumno en desarrollo de su trabajo final, Departamento de Informática, FCEFN - UNSJ
[1**]Alumno Becario de Investigación y en desarrollo de su trabajo final, FCEFN - UNSJ
[2]Gobierno de la Provincia de San Juan
[3]Instituto Tecnológico Buenos Aires
[4]Facultad de Arquitectura, Urbanismo y Diseño - UNSJ.

{mlund, silvina.migani}@unsj-cuim.edu.ar;{balmaceda.silvina, ferrarinicintia, juan.capdevila27, geralvarezn100, leodelosriosfacultad, danieltapella, leoblasco88}@gmail.com; ignacio.colome@unsj-cuim.edu.ar; ramiro@deployr.ai; {famascba, civerados, ericaminet}@gmail.com

RESUMEN

Hoy en día los datos constituyen el recurso más valioso para toda organización. Son el pilar fundamental para un eficiente proceso de toma de decisiones, imprescindible en esta acelerada economía digital. Sin embargo, el desarrollo de una solución de Analítica e Inteligencia Empresarial o de Negocio, presenta muchos y variados desafíos; desde enfrentar el procesamiento de una heterogénea y enorme cantidad de datos, hasta encarar problemas organizacionales. Es así que resulta fundamental contar con un proceso de desarrollo preciso y claro, además del conocimiento de estrategias, prácticas y herramientas apropiadas.

Dentro del ámbito del proyecto marco se han estudiado, desarrollado y/o aplicado procesos, y usado herramientas de software en diferentes dominios y alcances. Este trabajo expone algunas de las actividades desarrolladas.

Palabras claves: Toma de Decisión - Analítica e Inteligencia Empresarial - Procesos de Negocio - Almacén de Datos.

CONTEXTO

Este trabajo forma parte del proyecto PIC Código: E1212, correspondiente a la Convocatoria 2022 (Res. Nº 2902/22-R), denominado "*ANALÍTICA E INTELIGENCIA DE NEGOCIOS. DISEÑO DE UN MODELO DE PROCESOS DE NEGOCIO*" y del Acta Complementaria firmada entre la Secretaría de Tránsito y Transporte (STyT) del Gobierno de la Provincia de San Juan (GPSJ) y la UNSJ, aprobada por Resolución Nro. 113/2022-CD-FCEFN.

1. INTRODUCCIÓN

La Analítica e Inteligencia Empresarial o de Negocios (Analytics and Business Intelligence - ABI) tiene como propósito optimizar la

gestión de las organizaciones en lo referente al proceso de toma de decisiones, permitiendo mejorar o mantener su liderazgo en el mercado [1], [2].

Una característica esencial y común en este tipo de proyectos, es la necesidad de contar con la participación activa de los usuarios a lo largo de gran parte del proceso de desarrollo [3]. Sin embargo, tal como se mencionó en [4], son múltiples y variados los elementos, tecnologías y procesos que pueden utilizarse. La elección apropiada según cada caso, requiere un conocimiento profundo para poder generar una solución pertinente. Consecuentemente se investigaron y analizaron diferentes alternativas, como así también su aplicación en contextos diversos.

Principalmente se trabajó con datos relativos a pasajes de transporte público de colectivos de corta distancia de la provincia de San Juan (Acta firmada entre la STyT del GPSJ y la UNSJ) [5]. Aunque también se implementaron algunos procesos con datos del Ministerio de Educación del GPSJ y de una empresa privada dedicada a brindar servicios de Internet.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Las líneas de investigación que se abordan en el proyecto marco son:

- Identificación y definición de procesos que guíen en la generación de soluciones ABI.
- Análisis y estudio de tecnologías y herramientas útiles en la construcción de sistemas ABI.
- Aplicación de las tecnologías estudiadas y los procesos definidos en casos concretos.
- Análisis y diseño de interfaces interactivas adaptables a los usuarios.

Se ha trabajado en el estudio y uso de tecnologías y herramientas propias del almacenamiento y gestión de datos (PostgreSQL, MySql, Neo4j y Apache Spark), de la analítica descriptiva (Lenguaje SQL, PowerBI y Knowage) y también predictiva y prescriptiva (Pandas, Python y Jupyter Notebooks).

En cuanto a los datos, el proyecto se sustenta básicamente, con datos provenientes de la STyT del GPSJ, según acta antes mencionada. Asimismo se están realizando tareas con datos relativos a otros contextos: el Ministerio de Educación del GPSJ y una empresa privada proveedora de Internet. En ambos casos, corresponden a las actividades planteadas para el desarrollo de dos becas de investigación y de trabajos finales de licenciatura.

3. RESULTADOS

En el segundo año del proyecto, además de continuar con las líneas de investigación planteadas, se puso énfasis en divulgar, invitar y entusiasmar a alumnos de las carreras de informática, a participar en el proyecto. La gran cantidad de datos disponible brinda valiosas oportunidades de formación y crecimiento en temas de mucha actualidad como son: Analítica e Inteligencia de Negocios, Procesos de Negocio, Almacenes de Datos y Machine Learning. La convocatoria se focalizó sobre alumnos de 5to, 4to y 3er año (de estos últimos, sólo a alumnos que estaban cursando la asignatura Base de Datos). En las charlas iniciales participaron 24 alumnos, de los cuales mantuvieron el interés 12. Con el correr de los meses, cuando los alumnos iniciaron sus tareas de investigación, algunos comenzaron a delinear sus trabajos finales de licenciatura, becas y 4 debieron abandonar por compromisos laborales. Es así que se pidió la incorporación al proyecto de 8 alumnos. Por otra parte, también se integró al equipo de

trabajo, la Arquitecta Erica Minet Bravo, especialista en diseño y en procesamiento estadístico de datos.

El procesamiento de los datos ha presentado dificultades ante la falta de hardware adecuado. Además, las soluciones gratuitas en la nube presentan limitaciones.

A continuación se exponen algunas actividades de divulgación llevadas a cabo en el año 2024:

- Publicación del Trabajo "Analítica e Inteligencia Empresarial: Secretaría de Tránsito y Transporte de la Provincia de San Juan" en el XXVI Workshop de Investigadores en Ciencias de la Computación – XXVI WICC 2024 (UNPSJB, Puerto Madryn, Chubut, 18 y 19 de Abril). Autores: S.Migani, M.I.Lund, C.Ferrarini O., S.Balmaceda, E.Gil M., D.Tapella, L.Blasco, R.Savoie, F.Más, C.Vera, J.Capdevila, M.L.Molina, L.Drazic, M.Manassero. Libro de Actas: ISBN 978-987-8352-57-2. Libro de Pósteres: ISBN 978-987-8352-56-5 [4].

- Dictado de la Conferencia: "De la Gestión de la información a la Gestión del Conocimiento". Dentro del Congreso Internacional de Innovación en Sistemas, Tecnología e Información, realizado del 23 al 25 de Octubre de 2024, en el Instituto de Celaya, México. Conferencia Magistral dictada por M.Inés Lund y Silvina Migani [6].

- Divulgación en el marco del Día Internacional de la Mujer y la Niña en la Ciencia: "Negocios más inteligentes gracias a la tecnología". M.I.Lund y S.Migani [7].

FORMACIÓN DE RECURSOS HUMANOS

- Beca de Investigación y Creación categoría Estudiantes Avanzados. Tema: "Toma de decisiones apoyada en una solución de Inteligencia de Negocios. Caso de Aplicación: Ministerio de Educación de la provincia de San Juan", obtenida por el alumno Leonardo de los Rios, de la carrera LSI, a partir del 1 de diciembre del 2024. Res. N° 3182-R-2024.

- Beca de Investigación y Creación categoría Estudiantes Avanzados. Tema: "Integración de Datos en el contexto de la Analítica e Inteligencia Empresarial: Herramientas Computacionales aplicadas en una Empresa de Telecomunicaciones de San Juan", obtenida por alumno de la carrera LCC, Juan Manuel Capdevila, a partir del 1 de diciembre del 2024. Res. N° 3182-R-2024.

- Beca de Investigación y Creación categoría Estudiantes Avanzados. Tema: "Apache Spark en un Contexto de Inteligencia de Negocios: Análisis de Datos de Transporte Público de Pasajeros de Corta Distancia de la Provincia de San Juan.", obtenida por el alumno de la carrera LSI, Germán Álvarez Manzini, a partir del 1 de diciembre del 2024. Res. N° 3182-R-2024.

- Tesis de Maestría en Informática de la Lic. Cristina Vera, denominada "Neo4j como almacén de interacciones de GitHub: una evaluación empírica". En desarrollo. Dir: Silvina Migani.

- Tesis de Maestría en Inteligencia de Negocios y Explotación de Datos (Universidad Siglo XXI) del Lic.

Fernando Emanuel Más, denominada: "Implementación de un Selfservice de Datos para la gestión de KPIs en la Administración Tributaria". Presentada al tribunal para su defensa (marzo/abril del 2025). Dir: M.Inés Lund.

- Beca Manuel Belgrano del alumno Matías Ríos. Requisito: participar en proyecto de investigación durante el transcurso del año 2024.
- Trabajos Finales de Licenciatura en Sistemas de Información y en Ciencias de la Computación, en el marco del proyecto, de los alumnos:
  - Leonardo Drazic
  - Juan Manuel Capdevila
  - Leonardo de los Rios
  - Germán Alvarez Manzini
  - Ignacio Colomé

4. BIBLIOGRAFÍA

[1] J. Reis and M. Housley, *Fundamentals of Data Engineering: Plan and Build Robust Data Systems*. O'Reilly Media, 2022.

[2] R. Raj, S. H. S. Wong, and A. J. Beaumont, "Business Intelligence Solution for an SME: A Case Study," *Int. Jt. Conf. Knowl. Discov. Knowl. Eng. Knowl. Manag.*, vol. 3, pp. 41–50, 2016.

[3] S. Nalchigar and E. Yu, "Designing Business Analytics Solutions. A Moodel-Driven Approach," *Bus. Inf. Syst. Eng. 2018* , vol. 62, no. 1, pp. 61–75, Aug. 2018.

[4] S. Migani, M. I. Lund, and et al., "Analítica e Inteligencia Empresarial: Secretaría de Tránsito y Transporte de la Provincia de San Juan," in *XXVI Workshop de Investigadores en Ciencias de la Computación – XXVI WICC 2024*, 2024, pp. 194–198.

[5] "SUBE | Argentina.gob.ar." [Online]. Available: https://www.argentina.gob.ar/sube. [Accessed: 02-Mar-2024].

[6] M. I. Lund and S. Migani, "Conferencia: De la Gestión de la Información a la Gestión del Conocimiento," *CIISTI 2024 - Congreso Internacional de Innovación en Sistemas , Tecnología e Innovación*. Celaya - México, Oct-2024.

[7] M. I. Lund and S. Migani, "Negocios más inteligentes gracias a la tecnología." Fac. de Ciencias Exactas, Físicas y Naturales UNSJ, San Juan, Argentina, 2024.

Hibridación de Bases de Datos para el análisis de datos masivos

Graciela R. Etchart, Carlos E. Alvez, Silvia E. Ruiz, Ernesto M. Miranda, Juan J. Aguirre y Mariano R. Droz
Facultad de Ciencias de la Administración, Universidad Nacional de Entre Ríos (UNER),
{graciela.etchart, carlos.alvez, silvia.ruiz, ernesto.miranda, juan.aguirre, mariano.droz}@uner.edu.ar

Resumen

Actualmente, la diversidad en la organización de datos (estructurados, semi-estructurados y no estructurados) representa un desafío significativo para las empresas que necesitan maximizar el valor de su información. Las bases de datos relacionales (SQL) facilitan consultas rápidas y análisis precisos, mientras que las bases de datos NoSQL ofrecen flexibilidad y escalabilidad horizontal. En este trabajo se busca desarrollar una arquitectura genérica que combine las fortalezas de las bases de datos SQL y NoSQL para mejorar el análisis de grandes volúmenes de datos heterogéneos. Con este propósito se abordará la creación de un Sistema de Gestión de Bases de Datos Multimodelo (híbrido) que permita la integración efectiva de ambas tecnologías. La hibridación de estas tecnologías permitirá realizar análisis complejos y optimizar la toma de decisiones estratégicas, aprovechando la capacidad analítica de SQL y la flexibilidad de NoSQL

Palabras clave: NoSQL, Hibridación, Análisis, Big Data

Contexto

Este trabajo se da en el marco del proyecto de investigación "*Hibridación de bases de datos SQL y NoSQL para el análisis de datos masivos",* que da continuidad a los proyectos PID 07/G058*"Modelos de Machine Learning para la mejora de la precisión, seguridad y eficiencia en la gestión de datos biométricos"* y PID 07/044 "*Gestión de datos biométricos en base de datos objeto - relacionales*".

Introducción

El análisis de datos es esencial para las empresas y otros organismos, ya que permite convertir grandes volúmenes de datos en conocimientos prácticos y estratégicos. Las aplicaciones de análisis de datos son variadas y ofrecen numerosos beneficios, entre los que se incluyen la mejora en la toma de decisiones, la optimización de procesos y recursos, la generación de perfiles para mejorar la experiencia del cliente, la identificación de oportunidades y riesgos, entre otros [1].

Acceder a diversos tipos de organización de datos (estructurados, semi-estructurados y no estructurados) es crucial para el análisis, ya que las empresas deben optimizar la información disponible. Los datos estructurados, organizados en forma tabular, permiten realizar consultas rápidas y análisis precisos, como en el caso de las bases de datos relacionales [2]. Por otro lado, los datos semi-estructurados, como XML (eXtensible Markup Language) o JSON (JavaScript Object Notation), ofrecen una mayor flexibilidad al integrarse en diferentes sistemas [3]. Frecuentemente, los archivos XML y JSON se generan automáticamente y pueden ser utilizados para análisis. Por ejemplo, los sistemas ERP (Enterprise Resource Planning) o CRM (Customer Relationship Management) utilizan XML y JSON para intercambiar información, como facturas electrónicas u órdenes de compra; organizaciones del sector público o privado proporcionan datos estructurados en XML, como estadísticas o reportes meteorológicos; aplicaciones de comercio electrónico o redes sociales generan JSON para comunicar interacciones del usuario; y asistentes

virtuales como Alexa o Google Home utilizan JSON para comandos. Además, numerosos dispositivos IoT registran datos en formato JSON, útiles en sectores como la salud y el clima. En el ámbito agrícola, por ejemplo, los sensores de humedad, temperatura y calidad del suelo generan datos en JSON que luego se analizan para optimizar el riego y la fertilización. Por otro lado, los datos no estructurados, como textos, imágenes y videos, contienen información valiosa que, al ser analizada con herramientas avanzadas como la inteligencia artificial, revela patrones, tendencias y perspectivas estratégicas. Este enfoque integral potencia la toma de decisiones y fomenta la innovación.

La variedad de tipos de organización de datos es una de las cuestiones más desafiantes para el análisis de datos. En este sentido, en el entorno organizacional actual, la capacidad de tomar decisiones óptimas depende en gran medida de la habilidad para analizar y manipular datos con diferentes tipos de organización. Este desafío es aún mayor con la aparición de Big Data, que implica grandes volúmenes de datos de diversas fuentes y formatos que no pueden ser manejados eficientemente por sistemas tradicionales de gestión de bases de datos relacionales basados en el estándar SQL. Por lo tanto, se observa un aumento sustancial en la demanda de tecnologías y técnicas capaces de gestionar estos datos multimodelo, adaptándose a la naturaleza heterogénea y veloz de Big Data [4] [5].

Líneas de Investigación y Desarrollo

Una de las principales líneas de investigación de este trabajo es la hibridación de bases de datos SQL y NoSQL para el análisis de datos masivos.

No existe en la actualidad, un modelo de base de datos que se considere óptimo para la gestión de las diferentes organizaciones de datos, por el contrario, han surgido diferentes tipos de bases de datos acorde a cada caso. Esto es, Sistemas de Gestión de Bases de Datos (SGBD) que permiten gestionar de manera eficiente cada tipo de organización. Dado que estos nuevos SGBD no se rigen por el estándar SQL (para las bases de datos Relacionales y Objeto Relacionales), son comúnmente llamados NoSQL, que significa "No sólo SQL" (del inglés *not only SQL*) [6].

Los sistemas NoSQL no reemplazan a los sistemas de bases de datos con su lenguaje de consulta estándar SQL, ampliamente conocido y utilizado. Es importante destacar que las soluciones NoSQL no fueron creadas con los mismos propósitos que las soluciones basadas en SQL. De hecho, tanto las bases de datos SQL como las NoSQL tienen sus propias ventajas y desventajas [7]. Las bases de datos NoSQL, además de adaptarse a diferentes tipos de organización de datos, ofrecen los siguientes beneficios en comparación con las bases de datos relacionales:

- Escalabilidad más allá de las limitaciones de capacidad de los sistemas existentes: Las bases de datos SQL están diseñadas para organizar datos y permitir una escalabilidad vertical, lo que implica aumentar la capacidad de un único servidor (CPU, memoria, etc.) para manejar más datos o mayor carga. En cambio, las bases de datos NoSQL permiten la escalabilidad horizontal, es decir, agregar más servidores o nodos a un clúster distribuido para manejar más carga y datos, utilizando técnicas como el sharding (fragmentación de datos), la replicación o asignando diferentes datos a cada servidor.
- Agilidad y velocidad de desarrollo: Las organizaciones buscan adaptarse al mercado rápidamente y adoptar metodologías de desarrollo ágiles, DevOps y microservicios.
- Flexibilidad ante cambios: Las aplicaciones utilizan nuevos tipos de datos que cambian rápidamente. El modelo relacional, por lo general, es muy rígido ante cambios en las estructuras o tipos de datos. Las alternativas NoSQL se adaptan rápidamente al cambio y permiten gestionar datos estructurados, semi-

estructurados, no estructurados y polimórficos, así como manejar grandes volúmenes de estos datos.

Existen diferentes tipos de bases de datos NoSQL, las cuales buscan atender distintos requerimientos de aplicaciones. En este sentido, las bases de datos NoSQL se clasifican en cuatro categorías principales: modelo clave-valor, modelo de documento, modelo de grafos y modelo columnar [6]:

- Clave-valor: Cada elemento de la base de datos se almacena como un nombre de atributo o clave, junto con su valor. El valor es completamente opaco para el sistema y los datos sólo pueden ser consultados por la clave. Este modelo es útil para representar datos polimórficos y no estructurados, ya que no impone un esquema establecido en pares clave-valor. Es ideal para aplicaciones que solo consultan datos por un único valor clave, y ofrece ventajas en términos de rendimiento y escalabilidad debido a la simplicidad de los patrones de acceso y la opacidad de los datos. Ejemplos de SGBD: Redis y AWS DynamoDB.
- Documento: Almacenan datos en documentos con una estructura similar a JSON. A diferencia de las bases de datos relacionales, cada documento es un objeto, lo que simplifica el acceso a los datos y, en muchos casos, elimina la necesidad de operaciones JOIN y transacciones complejas de múltiples registros. Además, el esquema es dinámico y cada documento puede contener diferentes campos. Esta flexibilidad es muy útil para modelar datos no estructurados y polimórficos, y es beneficiosa para una amplia variedad de aplicaciones debido a la flexibilidad del modelo de datos, la capacidad de consulta y el mapeo natural de documentos a objetos. Ejemplos de SGBD: MongoDB y CouchDB.
- Grafos: Este modelo utiliza estructuras de grafos con nodos, enlaces y propiedades para representar los datos. Los datos se modelan como una red de relaciones entre elementos específicos, lo que facilita la navegación y el modelado de relaciones entre entidades en una aplicación. Es particularmente útil para aplicaciones donde las relaciones transversales son fundamentales, como la navegación por conexiones de redes sociales, topologías de redes o cadenas de suministro. Ejemplos de SGBD: Neo4j y AWS Neptune.
- Columnar: También conocido como modelo de columna ancha o familias de columnas, utiliza un mapa ordenado multidimensional distribuido y disperso para almacenar datos. A diferencia del modelo relacional, cada registro puede variar en el número de columnas almacenadas. Las columnas se pueden agrupar y distribuir en familias de columnas. Este modelo es ideal para aplicaciones con grandes volúmenes de datos que requieren alta disponibilidad, escalabilidad horizontal y lecturas rápidas de datos específicos, en lugar de realizar consultas complejas en varias columnas. Ejemplos de SGBD: HBase y Cassandra.

En este proyecto se trabajará en el área de SGBD multimodelo (híbrido) que provea una única plataforma de base de datos para gestionar datos SQL y NoSQL. Si bien las bases de datos multimodelo son un área emergente, en los últimos años los sistemas de bases de datos adoptan esta categoría [5]. En este contexto, la hibridación de bases de datos SQL y NoSQL es un tema de gran interés en la actualidad debido a la creciente necesidad de gestionar y analizar grandes volúmenes de datos heterogéneos [8].

Entre las implementaciones multimodelo prácticas se encuentran: *Netflix*, que utiliza Cassandra para gestión de datos en tiempo real y MySQL para métricas críticas; *Uber*, incluye PostgreSQL en su arquitectura para datos transaccionales y MongoDB para datos geoespaciales.

En este sentido, existen diversas propuestas para hibridar bases de datos SQL y NoSQL para aplicaciones específicas. En [9] se presenta un enfoque híbrido (NoSQL-SQL) para almacenar datos geoespaciales en MongoDB, estos datos son replicados y

mapeados en una base de datos PostgreSQL, utilizando una herramienta de código abierto llamada ToroDB Stampede. En [10] se realiza una revisión sobre tecnologías y técnicas para aplicaciones IoT. En tanto, en [11] se comparan distintas opciones de bases de datos SQL y NoSQL desde la perspectiva de la escalabilidad, para detectar falsos positivos en el reconocimiento biométrico de rostro.

En los trabajos citados anteriormente se trabaja con diferentes herramientas para la integración de bases de datos SQL y NoSQL. Las herramientas de integración son una parte importante para la creación de una arquitectura híbrida. Existen varias herramientas y técnicas que se pueden utilizar para integrar datos SQL y NoSQL para el análisis de Big Data. Herramientas como ODBC, JDBC, OData, REST y GraphQL permiten conectar y consultar orígenes de datos SQL y NoSQL. Los marcos de datos como Spark, Hadoop y Flink permiten el procesamiento y análisis de datos SQL y NoSQL utilizando un lenguaje o paradigma común. Los gestores de flujo de datos como Airflow, NiFi y Kafka facilitan la coordinación y automatización de la integración de datos SQL y NoSQL. Por último, los modelos de datos como ERD, JSON, XML y RDF proporcionan una representación lógica o conceptual para mapear y transformar datos SQL y NoSQL utilizando un esquema o estructura común [12].

No obstante, en los estudios revisados no se propone una arquitectura genérica para la integración enfocada en el análisis de datos. Por ello, este proyecto evaluará las posibilidades de integrar distintas bases de datos SQL y NoSQL utilizando herramientas disponibles en el mercado (principalmente de código abierto) para desarrollar una arquitectura híbrida genérica. Esta arquitectura debe ser capaz de gestionar interacciones masivas con datos heterogéneos provenientes de múltiples fuentes y tipos de organización, abarcando desde la fase inicial de recepción hasta la fase final de consulta, presentación y análisis, en diversos escenarios.

La hibridación de bases de datos SQL y NoSQL es un campo interdisciplinario que se nutre de una amplia variedad de teorías y modelos. Un sistema híbrido que combine bases de datos SQL y NoSQL se presenta como una solución óptima para el análisis de datos. En este contexto, para realizar análisis complejos que involucren grandes volúmenes de datos, es ventajoso combinar la capacidad de procesamiento analítico de las bases de datos SQL con la flexibilidad de las bases de datos NoSQL para gestionar datos de distintos tipos de organización.

Objetivos y resultados esperados

El objetivo general es proponer una arquitectura genérica para sistemas híbridos de bases de datos SQL y NoSQL para mejorar el análisis de grandes volúmenes de datos con diferentes tipos de organización, aprovechando las fortalezas de las bases de datos SQL y NoSQL.

Para cumplir con dicho propósito se revisarán los requisitos técnicos y las características de las bases de datos SQL y NoSQL, y se diseñará una arquitectura híbrida considerando aspectos como partición de datos, integración de esquemas y gestión de transacciones. Además, se propondrán métricas para evaluar el rendimiento de la arquitectura y se desarrollará un marco experimental para su implementación en diferentes escenarios.

Formación de Recursos Humanos

La estructura del equipo de investigación está formada por el Director, Co-Director e integrantes docentes.

Este trabajo servirá de marco para la tesis de Maestría en Sistemas de Información de la Facultad de Ciencias de la Administración de la UNER dirigida por el Director del Proyecto. Además, se propondrá como marco de trabajo para tesinas de grado de fin de carrera de Licenciatura en Sistemas de la mencionada

Facultad.
Por otra parte, se prevé incorporar un alumno de la carrera Licenciatura en Sistemas como Becario de Investigación.

Referencias

1. Mallam, P., Ashu, A., and Singh, B. 2021. Business Intelligence Techniques Using Data Analytics: An Overview. In Proceedings of the 2021 International Conference on Computing Sciences (ICCS), Phagwara, India, 2021, pp. 265-267. DOI: 10.1109/ICCS54944.2021.00059.
2. Elmqvist, N. 2023. Data Analytics Anywhere and Everywhere. Communications of the ACM, 66, 12 (Dec. 2023), 52-63. DOI: 10.1145/3584858.
3. Sadalage, P. J., and Fowler, M. 2013. NoSQL Distilled. A Brief Guide to the Emerging World of Polyglot Persistence. Addison-Wesley Professional.
4. Udeh, C. A., Orieno, O. H., Daraojimba, O. D., Ndubuisi, N. L., and Oriekhoe, O. I. 2024. BIG DATA ANALYTICS: A REVIEW OF ITS TRANSFORMATIVE ROLE IN MODERN BUSINESS INTELLIGENCE. Computer Science & IT Research Journal, 5, 1, 219-236. DOI: https://doi.org/10.51594/csitrj.v5i1.718.
5. Lu, J., and Holubová, I. 2019. Multi-model Databases: A New Journey to Handle the Variety of Data. ACM Comput. Surv., 52, 3, Article 55 (May 2020), 38 pages. DOI: https://doi.org/10.1145/3323214.
6. Kaufmann, M., and Meier, A. 2023. SQL and NoSQL Databases: Modeling, Languages, Security and Architectures for Big Data Management. Second Edition, Springer, June 2023.
7. Ali, A., Naeem, S., Anam, S., and Ahmed, M. M. 2023. A state of art survey for big data processing and nosql database architecture. International Journal of Computing and Digital Systems, 14, 1, 1-1.
8. Bjeladinovic, S., Marjanovic, Z., and Babarogic, S. 2020. A proposal of architecture for integration and uniform use of hybrid SQL/NoSQL database components. The Journal of Systems & Software. DOI: https://doi.org/10.1016/j.jss.2020.110633.
9. Herrera-Ramírez, J. A., Treviño-Villalobos, M., and Víquez-Acuña, L. 2021. Hybrid storage engine for geospatial data using NoSQL and SQL paradigms. Tecnología en Marcha, 34, 1 (Jan-Mar 2021), 40-54. DOI: https://doi.org/10.18845/tm.v34i1.4822.
10. Yadav, H. 2024. Structuring SQL/NoSQL databases for IoT data. International Journal of Machine Learning and Artificial Intelligence, 5, 5, 1-12.
11. Serengil, S. I., and Ozpinar, A. 2023. An evaluation of sql and nosql databases for facial recognition pipelines.
12. Youssef, S. E. A. 2023. Optimizing Database Performance for Large-Scale Enterprise Applications: A Comprehensive Study on Techniques, Challenges, and the Integration of SQL and NoSQL Databases in Modern Data Architectures. Journal of Artificial Intelligence and Machine Learning in Management, 7, 1 (2023), 81–92. DOI: https://journals.sagescience.org/index.php/jamm/article/view/166.

Indexación, Administración y Recuperación de Datos Multimedia Masivos

Luis Britos, Fernando Kasián, Verónica Ludueña, Diego Olivares,
Marcela Printista, Nora Reyes, Patricia Roggero, Gabriela Romano, Pablo Samat
LIDIC, Dpto. de Informática, Fac. de Cs. Físico Matemáticas y Naturales, Universidad Nacional de San Luis
{*fkasian, vlud, mprinti, nreyes, proggero*}@unsl.edu.ar,
{*ldoorz, olivarestello.diego, gabrielasindaromano, samatpablo*}@gmail.com
Karina Figueroa
Universidad Michoacana de San Nicolás de Hidalgo, México
karina@fismat.umich.mx
Claudia Deco
Facultad de Ciencias Exactas, Ingeniería y Agrimensura, Universidad Nacional de Rosario
deco@fceia.unr.edu.ar

Resumen

Los datos multimedia disponibles, provenientes de diversas fuentes digitales, se están almacenando y acumulando en grandes repositorios para tipos de datos no estructurados. Estos repositorios claramente no son modelizables bajo el modelo relacional de bases de datos, debido a la ausencia de estructura de los propios datos. Esto ha motivado que los índices aplicables y las herramientas que permiten administrarlos y realizar recuperación de información sobre ellos, utilicen nuevos modelos de representación, operaciones y nuevas maneras de realizar las búsquedas.

Esta necesidad de nuevas herramientas se hace más evidente si se considera el gran volumen de datos no estructurados que deben administrarse. Por ello, han surgido nuevos modelos y herramientas para su administración e indexación. Además, para resolver las operaciones de interés, considerando la existencia de la jerarquía de memorias y la posibilidad de hacer uso de computación de alto desempeño, estas nuevas herramientas deben ser eficientes en el uso de los recursos.

Palabras Claves: *bases de datos masivas, computación de alto desempeño, recuperación de información.*

1. Contexto

La línea "Recuperación de Datos e Información" pertenece al Proyecto Consolidado 03-1523 de la Universidad Nacional de San Luis (RCS 214/2023): "Tecnologías Avanzadas Aplicadas al Procesamiento de Datos Masivos" y se desarrolla en el Laboratorio de Investigación y Desarrollo en Inteligencia Computacional (LIDIC) de la UNSL. Se presentó en 2023 y estará vigente hasta 2026.

Esta línea cuenta con 4 docentes-investigadores y una estudiante de maestría y tiene como objetivo el desarrollo de herramientas eficientes para recuperar datos e información y administrar grandes bases de datos que almacenan datos no estructurados. Con este propósito se están desarrollando nuevas estructuras para indexar un gran volumen de datos no estructurados y se analizan nuevas técnicas que provean una buena interacción con el usuario. Estos índices deben resolver eficientemente las consultas de interés sobre distintos tipos de datos no estructurados; tales como: texto, secuencias de ADN, huellas digitales, audio, vídeo, etc.; deben ser escalables y estar especialmente diseñadas para memorias jerárquicas. Además, las técnicas de computación de alto desempeño (HPC) ayudan también a mejorar la eficiencia.

2. Introducción

En la actualidad se genera continuamente una gran cantidad de datos en formato digital, debido al uso masivo de internet, a la gran disponibilidad de dispositivos electrónicos y a la evolución de las tecnologías de información y comunicación. Por ello, estos datos digitales provienen de fuentes muy diversas y son además de tipos muy variados. Por lo tanto, se puede considerar en este problema que se evidencian dos características importantes de los problemas de "big data": el volumen de datos y la variedad de los mismos. Entonces, las consultas no pueden restringirse sólo a búsquedas por igualdad sobre datos estructurados tradicionales, porque de esa manera la visión completa del problema sería parcial Por lo tanto, se hace necesario redefinir las técnicas de

procesamiento, análisis y obtención de información útil y formular nuevas metodologías, para evitar que haya pérdida de información relevante debido a la elección del modelo.

Se habla de *búsqueda basada en contenidos* cuando lo que se utiliza es el dato no estructurado en sí mismo para describir lo que se está buscando. Este tipo de búsqueda es significativa para sistemas de recuperación de información sobre estos tipos de datos. Es obvio que no tiene sentido realizar, sobre estos tipos de datos, búsquedas que recuperen un objeto en la base de datos que sea idéntico al que se aporta. Así, se necesitan tipos de búsqueda más generales, tales como las *búsquedas por similitud*; y para resolverlas eficientemente, se utilizan los llamados índices o métodos de acceso *métricos* [3].

De esta manera, al considerar repositorios con grandes cantidades de datos no estructurados y querer responder consultas de recuperación de información, se debe recurrir a los índices métricos, porque ellos permiten lograr eficiencia en las respuestas. Además, cuando se piensa en aplicaciones reales sobre estos tipos de datos, los índices utilizados no sólo deben tener buen desempeño en las búsquedas, sino que también deben contar con operaciones de actualización del conjunto (altas y bajas de elementos). Otro aspecto importante es que estros índices deben ser *escalables*, porque al ser constante el crecimiento en el volumen de los datos, no debería afectarse su desempeño a medida que crece el volumen de datos a indexar.

En este escenario, un modelo adecuado es el de los *espacios métricos*, porque permiten modelizar las búsquedas por similitud. Formalmente, un espacio métrico se compone de un *universo* $\mathcal{U}$ de objetos y una *función de distancia* $d : \mathcal{U} \times \mathcal{U} \longrightarrow \mathbb{R}^+$. Dicha función mide la disimilitud entre dos objetos y cumple con las propiedades de reflexividad, positividad estricta, simetría y desigualdad triangular. Una de las principales ventajas de este modelo es que, además de brindar un marco teórico formal, es independiente del dominio de la aplicación.

Entre las variantes de las búsquedas por similitud, las más utilizadas son: la *búsqueda por rango* y la *búsqueda de los k vecinos más cercanos*. Si la base de datos $S \subseteq \mathcal{U}$, donde se realizan las búsquedas, posee n elementos ($\mid \mathcal{S} \mid = n$), cualquiera de estas búsquedas se puede resolver de manera trivial calculando n evaluaciones de distancia El cálculo de la función distancia puede ser muy costoso de calcular sobre algunos de estos tipos de datos no estructurados (por ej.: comparación de huellas digitales). Así, es importantes para lograr la eficiencia de los índices que se minimice la cantidad de cálculos de distancia que se realizan en las búsquedas durante el recorrido del índice. En general, si se conocen de antemano los elementos a indexar, se preprocesa la base de datos para construir el índice antes de comenzar a realizar búsquedas. En caso de no tener previamente los elementos de la base de datos S, se debe crear el índice de manera incremental; es decir, a medida que se conocen los elementos.

Más aún, en este contexto es probable que por el tamaño de la base de datos, del índice, o de ambos se deba utilizar como almacenamiento a la memoria secundaria. Pero, como los costos de acceso a memoria secundaria son más altos que a memoria principal, se debe minimizar además la cantidad de operaciones de E/S que se realizan durante una búsqueda. Por ello, para lograr eficiencia en las búsquedas no basta con realizar pocos cálculos de d sino que también se debe minimizar el número de operaciones de E/S. Así, claramente, en el diseño de los índices se debe considerar el nivel de la jerarquía de memorias sobre la que se trabaja. Además, se pueden utilizar técnicas de HPC y, en algunos casos, para responder más rápidamente a las consultas se pueden admitir respuestas no exactas [5].

Resumiendo,el objetivo de esta línea es desarrollar nuevas herramientas que soporten la interacción con el usuario, diseñar índices que permitan la manipulación eficiente de grandes volúmenes de datos no estructurados, para lograr recuperación de información eficiente sobre conjuntos masivos de este tipo de datos y contribuir al desarrollo de soluciones a problemas de big data para aplicaciones reales.

3. Líneas de Investigación

El objetivo es realizar aportes sobre distintos aspectos de los sistemas de recuperación de información en grandes volúmenes de datos no estructurados. Se han optimizado índices existentes, y diseñado nuevos índices, que permiten resolver distintas clases de consultas. Algunos de ellos se han incorporado en un sistema administrador de este tipo de bases de datos.

Métodos de Acceso Métricos

Las *bases de datos métricas* son un modelo adecuado para las bases de datos masivas que contienen datos no estructurados, porque allí tienen sentido las consultas por similitud. Además, como se pretende responder a estas consultas eficientemente, se deben

minimizar los costos de las búsquedas mediante el uso de los métodos de acceso métricos (MAMs) [3]. En general, durante las búsquedas se aprovechan las distancias almacenadas en el índice y los cálculos de distancia que se realicen durante el proceso, a fin de ahorrar cálculos de distancia, gracias a que la función de distancia cumple las propiedades de desigualdad triangular y simetría, porque se puede estimar la distancia entre cualquier objeto de consulta q y los elementos de la base de datos; para determinar qué objetos no son relevantes y evitar calcular d entre ellos y q. Cualquier ahorro en la cantidad de cálculos de distancia impacta en el tiempo necesario para responder a la consulta. Los MAMs almacenan algunas distancias precalculadas o referencias obtenidas al comparar los elementos de la base de datos respecto de un conjunto de objetos distinguidos llamados *centros*, *pivotes* o *permutantes*. Así, se clasifican en basados en particiones compactas, en pivotes o en permutaciones, respectivamente.

Existen varios aspectos a tener en cuenta al diseñar MAMs: el dinamismo (que soporten altas y bajas), en qué nivel de la jerarquía de memorias deben almacenarse, si es posible aplicar técnicas de HPC para mejorar los tiempos de respuesta, si son de respuesta exacta o de respuesta aproximada y la dimensionalidad del espacio métrico considerado.

Como ya se mencionó, por el volumen de los datos con los que se trabaja, los índices deben almacenarse en disco y se debe minimizar el número de cálculos de distancia y de operaciones de E/S en pos de la eficiencia. En esta situación, los índices sólo pueden disponer de la memoria principal para mantener información muy útil para determinar qué partes del índice deben recuperarse desde el disco.

Uno de los objetivos es el diseño de MAMs, preferentemente dinámicos y adaptados para memoria secundaria, con un buen desempeño en las búsquedas para las aplicaciones de interés. Se han desarrollado varios MAMs, algunos dinámicos y especialmente diseñados para grandes volúmenes de datos, basados en la *Lista de Clusters* (*LC*) [1] [2]. Uno de estos índices, que tiene una buena ocupación de página, operaciones eficientes tanto en cálculos de distancia como en operaciones de E/S y logra un buen desempeño en espacios de alta dimensión, es el *Conjunto Dinámico de Clusters* (*DSC*) [7]. El *DSC* organiza los centros de clusters en un *DSAT* en memoria principal y sus clusters en memoria secundaria [7]. La información en el *DSAT* es clave para disminuir la cantidad de cálculos de distancia y mantener bajos los costos de E/S.

En el *DSC* original, cada inserción debe ubicar el cluster donde se insertará el nuevo elemento, leerlo desde el disco y ver si el nuevo objeto podría ser un mejor centro para el cluster (se puede hacer más compacto el cluster usando ese elemento como centro); finalmente, se graba el cluster en el disco y se actualiza el *DSAT* de centros si el nuevo objeto se transformó en el nuevo centro. Como se puede observar, esta operación realiza los cálculos de distancia necesarios y pocas operaciones de E/S en el disco. Sin embargo, en las búsquedas, la calidad de los clusters obtenidos puede influir significativamente en el desempeño. Por ello, con vistas a mejorar la calidad de los clusters, se está analizando una variante que en lugar de insertar los elementos en el índice a medida que lleguen, demore la inserción de cada nuevo objeto a un cluster hasta tener varios elementos y determinar un mejor agrupamiento de los mismos. Así, se pueden mejorar los costos de búsqueda, al lograr clusters aún más compactos y asegurar una total ocupación de la página del disco, achicando el tamaño del archivo y reduciendo las operaciones de E/S y los tiempos de acceso. También permitirá disminuir el costo de construcción, amortizando el costo de grabar *un* cluster entre varias inserciones.

Por otra parte, el buen desempeño del *DSC* se debe a la calidad de la partición generada por los centros de sus clusters. Cada cluster está determinado por su centro y su radio, pero puede tener "huecos" dentro. Así, otro aspecto a considerar es mejorar la poda en estos agrupamientos, seleccionando un conjunto de "pivotes" globales que permitan caracterizar las zonas de los clusters dentro de las cuales existan efectivamente elementos [6]. Así, se mantiene para cada cluster información sobre la mínima y la máxima distancia a la que cada pivote encuentra elementos de dicho cluster. Esta información identifica los "huecos" del cluster. Durante las consultas, se puede descartar un cluster si la bola de la consulta no posee intersección efectiva con la zona del cluster que realmente contiene elementos de interés y se recuperan desde memoria secundaria todos los clusters cuya intersección con la bola de consulta con centro q y radio r es *no vacía*. Al identificar las zonas del cluster que no contienen elementos, si la consulta intersecta al cluster en su zona "hueca" se lo puede descartar y ahorrar cálculos de d y lecturas a disco. *DSCC* está en etapa de publicación.

[1] Se utiliza aquí la palabra *cluster* en inglés, en lugar de agrupamiento en espanõl, ya que es de uso habitual en computación.

Se ha diseñado e implementado el *Buffered On Line Dynamic Set of Clusters* (*BOLDSC*), una variante del *DSC*, que agrega un espacio en memoria para mantener los objetos que son insertados (bolsa), además de un índice auxiliar (de pivotes) sobre estos elementos [9]. Como pivotes se eligen elementos de la bolsa, y cuando ésta se llena, se selecciona el pivote que necesita el menor radio para encerrar la cantidad de elementos que caben en un cluster y todos éstos se sacan de la bolsa y se graban en un cluster completo. La grabación de clusters llenos mejora la cantidad de E/S, pero puede provocar que objetos cercanos, que se insertan en diferentes momentos, queden en clusters diferentes, ocasionando degradación en las búsquedas. Por ello, se está desarrollando una variante del *DSC* que graba clusters "casi llenos" (con cantidad de objetos en función de un factor de carga ρ). Al no llenarlos, habrá espacio libre en cada cluster y al insertar un nuevo objeto si hay clusters que lo puedan incorporar, se elige al más cercano a él entre ellos. Si este elemento es un mejor centro para el cluster, se actualiza el *DSAT* de centros. Si no es posible incorporarlo a un cluster (todos llenos), va a la bolsa y se actualiza el índice de pivotes. Así, los clusters se adaptan para mejorar los costos de búsqueda. Se ha finalizado su implementación y está en etapa de evaluación.

Como se ha mencionado,otra manera de acelerar las búsquedas es usar técnicas de HPC paralelizar el índice. Se ha tomado como base al *DSC* y se ha obtenido una versión paralela, que disminuye los cálculos de distancia y operaciones de E/S necesarias para responder a una consulta y resuelve más consultas simultáneas, aprovechando al máximo las operaciones de E/S realizadas. En esta implementación, cada procesador/nodo mantiene su propio índice local sobre sus datos locales y existe un broker, una cola de mensajería y los nodos conteniendo los índices locales. El broker se encarga de envíar los objetos al momento de crear el árbol y al hacer las búsquedas, dejando distintos mensajes en los tópicos correspondientes. Cada nodo lee los objetos del tópico de creación y los agrega al índice local. También lee los objetos de búsqueda del tópico, busca en el índice local y deja los resultados en ese tópico.

El broker y los nodos se comunican entre sí y los datos se almacenan como tópicos. En la creación, el broker deja un objeto en el tópico correspondiente, el nodo leerá de ese tópico e insertará el elemento en su índice local. En las consultas, el broker se encarga de leer las respuestas del tópico, unir los resultados, y devolver la respuesta que corresponde a cada consulta. Para identificar qué respuesta corresponde a qué consulta, se deja junto con el objeto un mensaje en el tópico con un identificador único. Así, la respuesta irá al tópico adecuado y el broker será capaz de unir todos los resultados del mismo identificador de tópico. Este índice se encuentra en etapa final de evaluación e implementación, para su próxima publicación.

El *Árbol de Aproximación Espacial Distal* (*DiSAT*) [1] es un índice estático, con buen desempeño en las búsquedas, debido a la buena partición inducida por el árbol sobre el espacio. Por ello, se puede obtener una *representación binaria compacta basada en esquemas*, utilizando la información de la partición inducida sobre la base de datos, y realizar búsquedas por similitud *aproximadas*. En estas búsquedas es posible intercambiar precisión en la respuesta por velocidad. Esta representación de secuencia de bits identifica la partición a la que pertenece el objeto y reduce el espacio de almacenamiento del índice. Durante las búsquedas, se obtiene la representación binaria de la consulta q y se la compara eficientemente con las representaciones de los objetos de la base de datos, para encontrar cuáles parecen ser similares a q (sus representaciones deben ser similares entre sí). Luego, la consulta q se debe comparar usando d para determinar si esos elementos son relevantes. Así, aunque podrían no recuperarse todos los objetos relevantes, se espera disminuir el costo de las búsquedas y obtener buena calidad de respuesta. En este caso, se encuentra avanzada la implementación del índice.

DBMS para Bases de Datos Multimedia

PostgreSQL es un DBMS que permite realizar consultas por similitud sobre algunos atributos. Por ejemplo, si se han creado índices *KNN-GiST* sobre texto, se pueden resolver búsquedas de k-vecinos más cercanos, pero no son aplicables sobre todo tipo de datos métricos. Además, estos índices brindan plantillas sólo para crear *árboles balanceados* que, tal como se ha mostrado en [2], no son adecuados para espacios métricos generales. Así, se debe incorporar a un DBMS la capacidad de manejar distintos tipos de datos no estructurados y sus operaciones.

Las operaciones habituales sobre bases de datos multimedia son las búsquedas por rango o de k-vecinos más cercanos. Sin embargo, en un DBMS para bases de datos multimedia [10], hay otras operaciones a considerar. Por ejemplo, se debe incorpo-

rar la operación de *join* por similitud [4]. Dadas dos bases de datos A y B, con $A, B \subseteq \mathcal{U}$, existen distintas variantes del join por similitud entre A y B. Estas variantes dependen del criterio de similitud Φ utilizado. Todas ellas devuelven un conjunto de pares (x, y), $x \in A$ e $y \in B$, tales que se satisface Φ entre x e y. Las variantes más conocidas son: el join por rango, el join de k-vecinos más cercanos y el join de k pares de vecinos más cercanos [8]. Formalmente, dadas $A, B \subseteq \mathcal{U}$, el *join por similitud* entre A y B es $(A \underset{\Phi}{\bowtie} B) = \{(x, y)/x \in A \wedge y \in B \wedge \Phi(x, y)\}$. Si $A = B$, la operación se denomina *auto-join* [4].

Una manera de acelerar los tiempos de respuesta de los joins consiste en disminuir los tamaños de los conjuntos obtenidos, a través de diversificar las respuestas [11]. De esta manera, el resultado es un conjunto más pequeño de pares, con respuestas útiles y diversificadas, y en menor tiempo. Permitir respuestas aproximadas a las consultas es otra manera de acelerar la respuesta. Se ha propuesto un algoritmo simple y eficiente para responder consultas de auto-join por similitud de k vecinos más cercanos aproximados, con una razonable precisión en la respuesta [4]. Con estos índices y algoritmos disponibles, se espera que *PostgreSQL* se transforme en un DBMS eficiente para datos multimedia masivos.

4. Resultados

Se han publicado los resultados de *BOLDSC* [9], se ha finalizado el diseño e implementación del *DSCC* y está en etapa de evaluación de los resultados, para su publicación.

Además, se está finalizando la evaluación de la nueva versión paralela del índice *DSC* y se espera también publicar estos resultados a la brevedad.

La extensión de *PostgreSQL*, para administrar y resolver consultas sobre bases de datos multimedia, pronto estará disponible, publicando los nuevos aportes al área.

5. Formación de Recursos

Se están realizando tesis de Maestría: (1) "Estructuras Eficientes sobre Datos Masivos para Búsquedas en Espacios" (UNSL), (2) "Sistema Administrador para Bases de Datos Métricas" (UNSL), (3) "Índices Métricos – Optimización del DSC usando Cortes de Regiones" (UNSJ), (4) "Optimización del BOLDSC por la Mejora de la Densidad y Solapamiento de los Clusters" (UNSJ) y (5) "Representación Basada en Esquemas para Búsquedas por Similitud Eficientes" (UNSL).

Referencias

[1] E. Chávez, V. Ludueña, N. Reyes, and P. Roggero. Faster proximity searching with the distal sat. *Inf. Systems*, 59:15 – 47, 2016.

[2] E. Chávez and G. Navarro. A compact space decomposition for effective metric indexing. *PRL*, 26(9):1363–1376, 2005.

[3] L. Chen, Y. Gao, X. Song, Z. Li, Y. Zhu, X. Miao, and C. S. Jensen. Indexing metric spaces for exact similarity search. *ACM Comput. Surv.*, 55(6), dec 2022.

[4] S. Ferrada, B. Bustos, and N. Reyes. An efficient algorithm for approximated self-similarity joins in metric spaces. *Inf. Systems*, 91:101510, 2020.

[5] K. Figueroa, A. Camarena-Ibarrola, R. Paredes, N. Reyes, and B. R. H. Martínez. Boosting the permutation index for similarity searching using an eclectic approach. In P. Pesado, editor, *Computer Science – CACIC 2022*, 159–174, Cham, 2023.

[6] J. Lokoč, J. Moško, P. Čech, and T. Skopal. On indexing metric spaces using cut-regions. *Inf. Systems*, 43:1–19, 2014.

[7] G. Navarro and N. Reyes. New dynamic metric indices for secondary memory. *Inf. Systems*, 59:48 – 78, 2016.

[8] R. Paredes and N. Reyes. Solving similarity joins and range queries in metric spaces with the list of twin clusters. *JDA*, 7:18–35, 2009.

[9] R. Paredes, N. Reyes, K. Figueroa, and M. Hoffhein. A new dynamic, secondary-memory metric index. In P. Pesado, P. Thomas, and F. Chichizola, editors, *Congreso Argentino de Cs. de la Computación (CACIC 2024)*, 520–529, Oct. 2024.

[10] C. Rong, C. Lin, Y. N. Silva, J. Wang, W. Lu, and X. Du. Fast and scalable distributed set similarity joins for big data analytics. In *IEEE 33rd International Conf. on Data Engineering (ICDE)*, 1059–1070, April 2017.

[11] L. Santos, L. Carvalho, W. Oliveira, A. Traina, and C. Jr. Traina. Diversity in similarity joins. In G. Amato, R. Connor, F. Falchi, and C. Gennaro, editors, *Similarity Search and Applications*, vol. 9371 of *LNCS*, 42–53. Springer, 2015.

Indexación, Búsquedas y Aplicaciones de Bases de Datos no Estructurados

Paola Azar[1], **Darío Ruano** [1], **Ezequiel Bernaldez**[1], **Ezequiel Nodar**[1], **Gerardo Ragazzi**[1], **Nahir Isray**[2], **Norma Herrera** [1]

[1]Laboratorio de Investigación y Desarrollo en Bases de Datos
Universidad Nacional de San Luis, San Luis, Argentina
dmruano@unsl.edu.ar, epazar@unsl.edu.ar, nherrera@unsl.edu.ar, ezebernaldez@gmail.com, ezequiel nodarhbk@gmail.com, gnragazzi@gmail.com

[2] Escuela de Ingeniería y Ciencias Ambientales
Universidad Nacional de Villa Mercedes, San Luis, Argentina
nahi24melani@gmail.com

Resumen

Los avances en las tecnologías de la información y la comunicación han permitido un crecimiento exponencial en el tamaño de las bases de datos, así como una mayor diversificación en su contenido. Un gran porcentaje de los datos disponibles son no estructurados, lo que dificulta su integración en el modelo relacional. Como respuesta a esta necesidad, han surgido nuevos modelos de datos y herramientas especializadas para su gestión. Nuestro interés se centra en el estudio de modelos para bases de datos no estructuradas, así como en el desarrollo de técnicas de indexación y búsqueda asociadas a estos modelos.

Contexto

El presente trabajo se desarrolla en el ámbito de la línea Técnicas de Indexación para Datos no Estructurados del Proyecto Tecnologías Avanzadas de Bases de Datos (PROICO 03-1723), cuyo objetivo es realizar investigación sobre manejo y recuperación eficiente de información no tradicional que implique el manejo de datos no estructurados. Este proyecto forma parte de Laboratorio de Investigación y Desarrollo en Bases de Datos (LaBDa) de la Universidad Nacional de San Luis y se encuentra dentro del Programa de Incentivos.

1 Introducción

El crecimiento exponencial de las fuentes de información disponibles, ha provocado que las bases de datos crezcan en volumen y en tipo de datos almacenados. En la actualidad, gran parte de la información disponible involucra el uso de datos no estructurados (sonidos, imágenes, video, huellas digitales, etc) que no admiten la representación clásica de bases de datos relacionales, donde la información se organiza como nuplas (filas) en relaciones (tablas). Los procesos de búsqueda enfrentan nuevos desafíos, tanto en términos de rendimiento como en la naturaleza de las respuestas esperadas.

En este contexto, han surgido nuevos modelos de bases de datos diseñados para cubrir

las necesidades de aplicaciones que manejan datos no estructurados. Entre ellos destacan el modelo de *bases de datos de texto* [4], el *modelo de espacios métricos* [7] y el modelo *métrico-temporal* [1]. En cada uno de estos modelos varía el tipo de consulta a resolver y, en consecuencia, el tipo de algoritmo de indexación y búsqueda a utilizar.

En el caso de las bases de datos textuales, la consulta más común consiste en recuperar texto a partir de una secuencia de caracteres, llamada patrón de búsqueda, ingresada por el usuario. Si la base de datos almacena texto en lenguaje natural, la recuperación de documentos se realiza a partir de palabras claves proporcionadas por el usuario.

En el caso de los espacios métricos, las consultas se basan en la similitud entre elementos. Este tipo de consultas, conocidas como *query by example*, consisten en proporcionar un elemento como referencia para recuperar otros similares. Más específicamente, el usuario ingresa un elemento, que puede o no formar parte de la base de datos, y el sistema rettorna los objetos más similares dentro de ella.

Finalmente, en el modelo métrico temporal, las búsquedas también se basan en la similitud entre elementos, pero incorporando la variable tiempo. En este caso, cada elemento de la base de datos tiene un intervalo de vigencia asociado. Las búsqueda tienen como objetivo recuperar objetos similares a uno dado, pero restringiendo la búsqueda a un rango de tiempo específicado por el usuario.

En todos estos modelos, para resolver las consultas de manera eficiente, es fundamental el uso de índices diseñados específicamente para cada tipo de dato y de consulta planteada.

2 Líneas de Investigación

En este proyecto nos enfocamos en el estudio, diseño y optimización de algoritmos de indexación y búsqueda para datos no estructurados, además de explorar sus posibles aplicaciones en casos reales. La implementación de las técnicas desarrolladas en casos reales nos permite transferir tecnología al medio y evaluar el desempeño de nuestras propuestas en problemas concretos de la comunidad. En este sentido estamos trabajando en problemas relacionados a medicina forense.

Describimos a continuación las líneas de investigación abordadas.

2.1 Indexación y Búsquedas

Una de las líneas de trabajo abordadas es el diseño de índices para memoria secundaria, lo que implica desarrollar estrategias eficientes de paginación. El proceso de paginación de un índice consiste en dividirlo en partes, cada una de las cuales se almacena en una página de disco. Durante la búsqueda, se carga una parte en la memoria principal, se realiza la consulta en ella y se repite el proceso hasta completar la búsqueda.

Cuando un índice se maneja en disco, el costo de búsqueda queda determinado por la cantidad de accesos a disco realizadas [5]. Por ello, el proceso de paginación debe garantizar que solo se acceda a las páginas que contienen información relevante para la consulta.

Actualmente, estamos trabajando en la implementación de dos índices para memoria secundaria: el Trie de Sufijos (TS) [3], utilizado en bases de datos textuales , y el Historical-FHQT (H-FHQT) [2], diseñado para bases de datos métrico-temporales. Hasta el momento, hemos desarrollado e implementado versiones funcionales del TS y del H-FHQT en memoria secundaria. Actualmente, ambos índices se encuentran en fase de evaluación experimental.

2.2 Aplicaciones a Problemas de Medicina Forense

En el ámbito de la medicina legal y forense existen múltiples problemas de relevancia que necesitan del soporte de herramientas

informáticas para su resolución. En este proyecto hemos abordado dos de ellos: identificación de cadáveres NN y datación de la muerte.

Identificación de Cadáveres

Dentro de los cuerpos que ingresan a los distintos Institutos de Medicina Forense del país, existen casos que no poseen las condiciones adecuadas para su identificación inmediata (indocumentados, en avanzado estado de descomposición, restos óseos, etc.) o sin posibilidad de identificación (fragmentos muy pequeños, restos carbonizados, etc.). Frente a esto, las instituciones deben investigar no sólo para determinar qué fue lo que sucedió (causa de la muerte) y cuándo sucedió (data de la muerte), sino también para poder dar con la identidad del cuerpo. Claramente la identificación de cadáveres está directamente relacionada con la búsqueda de personas desaparecidas.

Hemos abordado el estudio de esta problemática con el objetivo de desarrollar un sistema que permita gestionar una base de datos que colabore en identificación de cadáveres NN en el contexto de personas desaparecida.

Usamos el modelo de espacios métricos para el diseño de la base de datos. Un espacio métrico está formado por un conjunto de elementos X y una función de distancia d que mide la similitud entre elementos de X.

A cada cadáver se le asocia un vector que representa sus características físicas (vector característico). El conjunto de vectores característicos de todos los cadáveres registrados conforman el conjunto X. Para el vector característico solo se consideran los datos que realmente sean importantes a la problemática de identificación. Usar pocos datos podría provocar que las búsquedas que se realicen sean de muy baja selectividad y usar demasiados puede provocar que se descarten elementos de la base de datos que sean de interés. En este sentido, ya hemos definido un primer core de datos sobre el cual trabajar.

Al realizar una búsqueda, es necesario ingresar las características físicas de la persona buscada para generar su vector característico. A partir de este vector, se ejecuta una búsqueda por similitud en la base de datos de cadáveres, obteniendo como resultado una lista de posibles coincidencias. Estos resultados se presentan ordenados según el grado de similitud con la persona buscada, facilitando así el proceso de identificación.

Para medir la similitud entre vectores característicos, usamos usaremos como función de distancia d la función coseno [7].

Un aspecto clave en este proceso es la asignación de pesos a las características, ya que no todas tienen el mismo grado de importancia en la identificación. Por ejemplo, el color de cabello es menos relevante que el color de ojos, ya que una persona puede cambiar el primero, pero no el segundo. Por ello, es fundamental definir un peso para cada característica que refleje su importancia relativa en el proceso de comparación.

Otro aspecto relevante es que, para algunas componentes del vector característico, no se cumple la propiedad de simetría, indispensable en cualquier función de distancia. Por ejemplo, si la persona buscada tiene una amputación en un brazo y el cadáver con el que se lo compara no presenta dicha amputación, la coincidencia debe descartarse, sin importar el grado de similitud en otras características. Sin embargo, si la persona buscada no tiene amputaciones, pero el cadáver sí, no debe descartarse ya que la amputación pudo haber ocurrido después de su desaparición. Este tipo de consideraciones son fundamentale para garantizar el resultado de la búsqueda.

Con respecto a la indexación, comenzamos utilizando algoritmos basados en pivotes. Una vez que la base de datos se alimente con datos reales, se podrá analizar la dimensionalidad del espacio métrico en el que se trabaja y, si es necesario, ajustar o reemplazar el algoritmo de indexación para optimizar el rendimiento.

En lo que respecta al tipo de búsqueda por similitud a usar, estamos explorando tres estrategias: búsqueda de los k vecinos más cercanos, búsqueda por rango usando un radio de búsqueda r y búsqueda dentro de clusters de vectores (ver sección 2.3). Para las dos primeras, estamos determinando experimentalmente los valores óptimos de k y del radio de búsqueda r.

Datación de la Muerte

El cronotanatodiagnóstico (o datación de la muerte) es el conjunto de observaciones y técnicas que permiten señalar el intervalo de tiempo donde con mayor probablidad se ha producido el proceso de muerte. Este intervalo, conocido como IPM (intervalo post mortem), indica un período de tiempo circunscripto por el momento de muerte y el hallazgo del cadáver. Conseguir averiguar este intervalo de forma precisa es de suma importancia tanto en el ámbito penal como civil.

La estimación del IPM es una de las tareas más complicadas en medicina legal, no existiendo un método que sea completamente exacto para ello. El apoyo de herramientas informáticas con las que cuenta un médico forense de nuestro país para datar la muerte es escaso o nulo en algunos casos. El software existente está basado en otras realidades: piden datos que en nuestro país no son viables y/o utilizan en sus cálculos fórmulas basadas en condiciones climáticas poco probables de ocurrir en nuestro país. Para complejizar aun mas la situación, no existe un núcleo común de datos que sea utilizado en todas las provincias para datar la muerte.

En este Proyecto abordamos este problema con el objetivo es diseñar un sistema de base de datos adecuado para la comunidad tanatológica argentina, que permita no solo realizar la datación de la muerte sino que también permita encontrar casos similares al que se está tratando. Esto último ha sido marcado como un punto de sumo interés en el proceso de datación por la comunidad forense.

Hasta el momento hemos desarrollado el prototipo del mismo y estamos en la etapa de estudio de algoritmos de clustering aplicables en este contexto.

2.3 Algoritmos de Clustering

Como mencionamos, en los problemas de medicina forense que estamos estudiando es de suma importancia la detección de clusters. El clustering es una técnica de aprendizaje no supervisado que permite agrupar datos según sus similitudes.

En el caso de indetificación de cadáveres, el clustering facilitará la organización de la información disponible sobre las personas NN en grupos con características similares. Esto permitirá reducir el conjunto de datos relevantes en cada búsqueda, optimizando los tiempos de respuesta y mejorando la precisión de los resultados.

Para la datación de la muerte, el conocer casos similares ayuda al médico forense a dar una estimación mas precisa del IPM. En este contexto, se dispone de un conjunto de casos en los que el IPM es conocido. Ante un nuevo caso, se deben identificar casos similares de datación para que el médico forense pueda comparar y corrobar sus resultados, mejorando la precisión de su análisis.

Hemos considerado el estudio de dos algoritmos de clustering:

K-Means: un método eficiente que agrupa los datos en un número predefinido de grupos. Sin embargo, requiere definir el número de grupos de antemano, lo que puede ser una limitación si la cantidad óptima de clústeres no es conocida previamente.

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) [6]: adecuado cuando existen agrupaciones de diferentes densidades, perimitiendo ademá identificar datos atípicos (outliers), lo que podría ser

beneficioso en casos de individuos con características poco comunes.

Para el problema de datación, estudiaremos el comportamiento de ambos algoritmos.

Para el problema de identificación de cadáveres, consideramos que el algoritmo más adecuado es DBSCAN ya que no requiere un número predefinido de clusters. Además, DBSCAN es capaz de manejar tanto distribuciones densas como dispersas, una característica clave para trabajar con datos biológicos, que no suelen seguir una distribución homogénea. Otra ventaja importante es su capacidad para identificar casos atípicos, lo que permite aislar individuos con características poco comunes, facilitando así su identificación.

3 Resultados Esperados

Como resultados de los trabajos que estamos realizando, esperamos obtener índices eficientes en memoria secundaria para las bases de datos planteadas. Todo el trabajo realizado nos permite adquirir la experiencia suficiente como para abocarnos al diseño de otros índices que sean competitivos en este ámbito.

En cuanto a la transferencia tecnológica, el resultado esperado es un sistema web de identificación de cadáver en el contexto de búsqueda de personas desaparecidas y un sistema de soporte para la datación de la muerte. En ambos casos contamos ya con los prototipos funcionando.

4 Formación de Recursos Humanos

Dentro de esta línea se forman docentes y alumnos de la Universidad Nacional de San Luis y de la Universidad Nacional de Villa Merecedes. Tres alumnos de la Licenciatura en Ciencias de la Computación están realizando una pasantía de investigación en este Proyecto, colaborando con los trabajos aquí expuestos. Actualmente hay en desarrollo 2 Tesis de Maestría y un Trabajo Final de la Licenciatura, todos de la Universidad Nacional de San Luis.

Bibliografia

[1] A. De Battista, A. Pascal, N. Herrera, and G. Gutierrez. Metric-temporal access methods. *Journal of Computer Science & Technology,*, 10(2):54–60, 2010.

[2] A. De Battista, A. Pascal, G. Gutierrez, and N. Herrera. Un nuevo índice métrico-temporal: el historical-fhqt. In *Actas del XIII Congreso Argentino de Ciencias de la Computación*, Corrientes, Agentina, 2007.

[3] G. H. Gonnet, R. Baeza-Yates, and T. Snider. *New indices for text: PAT trees and PAT arrays*. Prentice Hall, New Jersey, 1992.

[4] Navarro. *Text Databases*, pages 688–694. Idea Group Inc., Pennsylvania, USA, 2005.

[5] Jeffrey Scott Vitter. *Algorithms and Data Structures for External Memory*. Publishers Inc, 2006.

[6] Wei-Tung Wang, Yi-Leh Wu, Cheng-Yuan Tang, and Maw-Kae Hor. Adaptive density-based spatial clustering of applications with noise (dbscan) according to data. In *2015 International conference on machine learning and cybernetics (ICMLC)*, volume 1, pages 445–451. IEEE, 2015.

[7] Pavel Zezula, Giuseppe Amato, Vlastislav Dohnal, and Michal Batko. *Similarity Search: The Metric Space Approach*. Springer Publishing Company, Incorporated, 1st edition, 2010.

Líneas de investigación complementarias en el marco de proyectos de Aplicación de Tecnologías de Inteligencia Artificial

Osvaldo Sposito[1], Luis Busnelli[2], Julio Bossero[1], Viviana Ledesma[1], Lorena Matteo[1], Sebastián Quevedo[1], Carlos Colombain[1], Laura Conti[2], Gastón Procopio[1], Celia Gargano[1], Gerardo Frega[2], Sergio García[3].

[1]Universidad Nacional de La Matanza. Departamento de Ingeniería e Investigación Tecnológicas. {sposito, jbossero, vledesma, gprocopio, lmatteo, cgargano, ccolombain, squevedo}@unlam.edu.ar
[2] Universidad Nacional de La Matanza. Departamento Derecho y Ciencia Política.{ lbusnelli, lconti, gfrea}@unlam.edu.ar
[3] Palacio de Tribunales. Departamento Judicial de Morón. Alte. Brown. Piso 4. Morón. sergiogabriel.garcia@pjba.gov.ar

RESUMEN

Este trabajo describe dos líneas de investigación convergentes que abordan desafíos actuales en inteligencia artificial y recuperación de información, orientadas hacia la mejora de la toma de decisiones y la optimización de sistemas digitales. Por un lado, se exploran sistemas expertos híbridos que integran redes neuronales artificiales y minería de datos para soportar procesos de decisión en el ámbito judicial, que podrían aumentar la precisión y adaptabilidad de los sistemas existentes. Por otro, se trabaja en la mejora de algoritmos de clustering, como K-Means, y técnicas avanzadas de procesamiento del lenguaje natural para la recuperación eficiente de documentos digitalizados en contextos administrativos. Estas líneas representan un esfuerzo continuo por parte del equipo para desarrollar soluciones innovadoras que combinan técnicas de inteligencia artificial y ciencia de datos, con aplicaciones en los ámbitos judicial y administrativo.

Palabras clave: *Inteligencia Artificial, Recuperación de Información, Sistemas Expertos, Redes Neuronales, Procesamiento del Lenguaje Natural, K-Means.*

CONTEXTO

El presente trabajo se desarrolla en el marco de proyectos de investigación que abordan el uso de tecnologías avanzadas para la optimización de sistemas de decisión y recuperación de información, específicamente: *Justicia Tecnológica: Reforzando Experticia con Redes Neuronales para una toma de decisiones más eficiente* (C262/PII - 2024-2025) y *Optimización de un Sistema de Recuperación de Información para Expedientes Digitalizados en la UNLaM: Integración de Algoritmo K-Means Mejorado y Exploración de Técnicas Avanzadas* (C268 - 2025-2026).

Ambos proyectos pertenecientes al Programa de Incentivos para Docentes Investigadores de la Secretaría de Políticas Universitarias del Ministerio de Educación de la Nación. Los integrantes de este trabajo son investigadores y docentes dependientes de las siguientes unidades académicas de la Universidad Nacional de La Matanza (UNLaM): el Departamento de Ingeniería e Investigaciones Tecnológicas, y el Departamento Derecho y Ciencia

Política, además, colaboran personal técnico de la Subsecretaría de Tecnología Informática del Poder Judicial de la Provincia de Buenos Aires.

1. INTRODUCCIÓN

El Proyecto C262/PII es una continuación de los proyectos previos C236/PII y C249/PII [1-2] y se enfoca en optimizar el proceso de razonamiento del sistema Experticia [3-5], una herramienta desarrollada por este equipo en investigaciones anteriores para brindar soporte en la toma de decisiones judiciales. Si bien Experticia ha demostrado ser útil en los proyectos previos, se han identificado áreas de mejora en su desempeño específicamente mediante la incorporación de Redes Neuronales Artificiales (RNA) y algoritmos de Minería de Datos [6-7]. El objetivo principal de este proyecto es evaluar si la implementación de RNA puede mejorar la toma de decisiones judiciales, comparando su eficacia con los actuales árboles de decisión utilizados en Experticia.

Diversos estudios han explorado estrategias para mejorar los "Sistemas Expertos Híbridos", combinando técnicas de Inteligencia Artificial (IA), como RNA, con el objetivo de incrementar su eficiencia y precisión. La literatura consultada sugiere que la integración de RNA en los sistemas expertos potencia su capacidad de aprendizaje y adaptación, aprovechando tanto el procesamiento de patrones complejos como la representación del conocimiento y el razonamiento lógico [8-11]. Esta combinación permite que el sistema evolucione a partir de datos de entrada, facilitando la mejora continua y la adaptación a nuevos escenarios en el ámbito jurídico.

En cuanto al proyecto C268, es importante destacar que la Comisión de Conectividad y Sistemas de Información del Consejo Interuniversitario Nacional (CIN), elaboró un informe en 2018 a través de la Subcomisión Técnica del Sistema de Información Universitaria (SIU), destacó la necesidad de implementar un sistema de expediente electrónico en el ámbito universitario argentino [12]. Para llevar adelante esta iniciativa se seleccionó un sistema desarrollado por la Universidad Nacional de General Sarmiento (UNGS). Este desarrollo denominado SUDOCU (Sistema Único Documental), es un sistema de gestión documental integral que permite administrar documentos, expedientes y trámites.

La UNLaM implementó el SUDOCU durante en 2024. Por otra parte, ha avanzado en la digitalización del histórico de documentos, expedientes y trámites. A medida que aumenta el volumen de información, la eficiencia en la recuperación de datos se vuelve crucial. Sin embargo, el sistema actual que administra el histórico no optimiza la búsqueda basada en el contexto semántico de los documentos. Implementar técnicas de Recuperación de Información, como el modelo de espacio vectorial o modelos avanzados de aprendizaje profundo y procesamiento de lenguaje natural, podría mejorar la precisión y utilidad de las búsquedas, especialmente en contextos técnicos o ambiguos.

2. LÍNEA de INVESTIGACIÓN y DESARROLLO

El equipo de investigación ha venido trabajando en dos líneas principales que integran técnicas de IA y recuperación de información para abordar desafíos actuales en diversos ámbitos:

- **Integración de IA en sistemas expertos**
  - Desarrollo y evaluación de sistemas híbridos que combinen RNA con sistemas expertos

tradicionales para mejorar la precisión y adaptabilidad en la toma de decisiones.
- Exploración de arquitecturas avanzadas de RNA, para el análisis y procesamiento de patrones en datos judiciales.
- Estudio de algoritmos de aprendizaje supervisado para optimizar la interpretación de casos judiciales y generar predicciones más confiables.

 Optimización de Algoritmos de Recuperación de Información
- Desarrollo de técnicas de análisis semántico latente y modelos de espacio vectorial para mejorar la búsqueda y clasificación de documentos en sistemas digitales.
- Modificación y evaluación de algoritmos de clustering, como K-Means, para minimizar cálculos redundantes y aumentar la eficiencia sin comprometer la calidad de los resultados.
- Aplicación de técnicas de procesamiento del lenguaje natural, incluyendo lematización y descomposición en valores singulares, para mejorar la precisión semántica en sistemas de búsqueda contextual.

3. RESULTADOS OBTENIDOS/ESPERADOS

En cuanto al proyecto C262/PII, que se encuentra en una etapa inicial de definición y diseño. En primer lugar, se realizó una búsqueda bibliográfica de los distintos sistemas expertos híbridos. Este trabajo se presentó en el Congreso Nacional de Ingeniería Informática / Sistemas de Información (CoNaIISI) 2024. Bajo el título "*Optimización de Experticia: Revisión de Sistemas Híbridos con Redes Neuronales y Sistemas Expertos*". El mismo revisa la literatura sobre estos sistemas durante el período abarcado entre 2019 y 2024. Se han analizado diversos enfoques, incluyendo redes neuronales convolucionales, perceptrones multicapa y redes profundas, que han demostrado su eficacia en áreas específicas como también en la optimización de sistemas expertos. Los resultados indican que estas tecnologías pueden mejorar la precisión y la interpretabilidad de los sistemas de decisión.

En relación con el proyecto C268, el equipo ha avanzado en el análisis y mejora del algoritmo K-Means, enfocándose en optimizar su eficiencia computacional. En CACIC 2024 [13], se presentó un trabajo donde se evaluó una propuesta de optimización diseñada para reducir el número de cálculos de distancia entre puntos y centroides, manteniendo la calidad del agrupamiento.

En el K-Means estándar, cada iteración implica calcular la distancia de cada punto a todos los centroides para reasignarlo al clúster más cercano. Sin embargo, muchos puntos permanecen en el mismo clúster sin cambios, lo que genera cálculos innecesarios. Para mitigar esta redundancia, se implementaron las siguientes optimizaciones:

- **Zonas fijas dentro de los clústeres**

✓ En cada iteración, se identificaron puntos que en iteraciones previas no habían cambiado de clúster y cuya distancia al centroide estaba dentro de un umbral de estabilidad.

✓ Estos puntos fueron marcados como "fijos" y sus distancias no fueron recalculadas en las siguientes iteraciones, a menos que el centroide se moviera significativamente.

- **Limitación de cálculos basados en proximidad relativa**

✓ En lugar de calcular la distancia de cada punto a todos los centroides, se restringió la evaluación solo a

aquellos centroides ubicados dentro de un radio determinado.

- ✓ Para definir este radio, se utilizó la distancia del punto al centroide más cercano en la iteración previa y un umbral de cambio de centroides observado en iteraciones anteriores.

Estas modificaciones se implementaron en un prototipo que simuló el proceso de agrupamiento en un corpus de documentos, utilizando representaciones vectoriales de términos con ponderaciones TF-IDF.

Los experimentos mostraron que la optimización permitió reducir el número de cálculos de distancia en un 81.3 %, pasando de 39000 a solo 7303 cálculos, sin afectar la calidad del agrupamiento. Como se observa en la Tabla 1, el coeficiente de silueta se mantuvo inalterado en comparación con K-Means estándar, y el número de iteraciones necesarias para la convergencia no cambió.

Tabla 1. Comparación entre el K-Means original y el optimizado.

Medida	Original	Optimización	Comparación
Silueta	-0.24	-0.24	100%
Iteraciones	13	13	100%
Cantidad Cálculos	39000	7303	18,72%

En la Figura 1, se observa el impacto de la optimización en la distribución de los clústeres. Los centroides (representados por diamantes) y los puntos asignados a cada clúster (círculos) se diferencian por color. Las regiones sombreadas indican áreas donde los cálculos pueden reducirse en iteraciones posteriores. En particular, la zona fija del clúster, destacada en tonos morados, representa los puntos cuya asignación permanece estable, evitando cómputos innecesarios en las siguientes iteraciones.

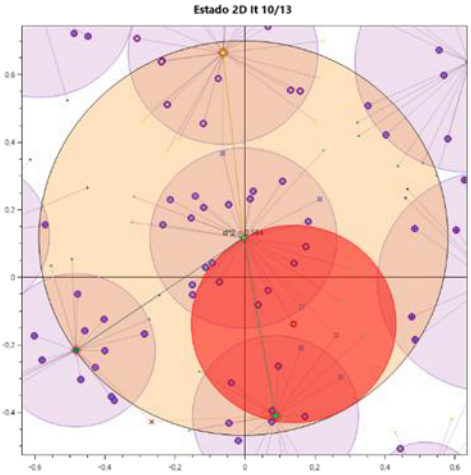


Figura 1 Visualización de los clústeres con K-Means mejorado. Fuente: Tomado de [13]

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo está compuesto por un total de 17 integrantes, distribuidos en los dos proyectos mencionados anteriormente, cada uno aportando su experiencia y conocimientos al desarrollo de la línea de investigación. Entre ellos se destacan:

- Dos integrantes finalizaron el doctorado en Ciencias Jurídicas.
- Un miembro dedicado actualmente a la elaboración de su tesis doctoral.
- Dos integrantes que han finalizado sus estudios de maestría en Informática en la UNLaM.
- Cuatro integrantes incorporados en la "Carrera de Docente Investigador UNLaM" en la categoría de Investigador en Formación.

Se prevé la realización de cursos de actualización y capacitación en el área de estudio, con el objetivo de fortalecer las habilidades y conocimientos del equipo en aspectos clave para la investigación. Además, se espera incorporar al equipo alumnos que consoliden su formación como futuros investigadores, contribuyendo al desarrollo continuo de estas líneas de trabajo.

5. BIBLIOGRAFÍA

[1] C236/II – Proyecto de investigación PROINCE.” Diseño e Implementación de un

Sistema Experto como Apoyo al Proceso de Despacho de Trámites de un Organismo Judicial". Director: Osvaldo Spositto. Período 2020-2021.

[2] C249/ PII – Proyecto de investigación PROINCE. "Inteligencia Artificial Jurídica: la Evolución de Experticia hacia un Modelo de Justicia Predictiva". Director: Osvaldo Spositto. Período 2022-2023.

[3] Sistema Experto para Apoyo del Proceso de Despacho de Trámites de un Organismo Judicial. Spositto, O.; Busnelli, L.; Conti, L.; García, S.; Pérez Villar, G.; Ledesma, V.; Procopio, G.; Bossero, J. XIV Simposio de Informática en el Estado (SIE 2020) - JAIIO 49. Facultad de Ingeniería de la UBA. ISSN: 2451-7534, pp. 17-29. Octubre 2020.

[4] Metodología para evaluar un modelo de Justicia Predictiva. Spositto, O.; Busnelli, L.; Conti, L.; García, S.; Pérez Villar, G.; Ledesma, V.; Procopio, G.; Bossero, J. CoNaIISI 2020. UTN – Fac. Reg. San Francisco. ISBN 978-950-42-0202-8, pp. 527-535. Noviembre 2020.

[5]Experticia. Un Modelo de Sistema Experto aplicada al Poder Judicial. Spositto, O.; Busnelli, L.; Ledesma, V.; Conti, L.: García, S.; Procopio, G.; Pérez Villar, G.; Frega, G.; Bossero, J.; Saizar V.; Quintana, F. XXIII. WICC 2021. UN C, La Rioja. ISBN: 978-987-24611-3-3; 978-987-24611-4-0, pp. 113-118. Abril 2021.

[6] José Ignacio Solar Cayón, (2020). "La inteligencia artificial jurídica: nuevas herramientas y perspectivas metodológicas para el jurista", Revus [Online], 41, Disponible en: http://journals.openedition.org/revus/6547, Fecha de consulta: 19/02/2025

[7] Pérez, Hugo. (2023). "Utilización de la inteligencia artificial de sistema experto en la investigación del delito". Revista Mexicana de Ciencias Penales. 7. 3-20. 10.57042/rmcp. v7i21.674. Disponible en: https://www.researchgate.net/publication/374384976_Utilizacion_de_la_inteligencia_artificial_de_sistema_experto_en_la_investigacion_del_delito/citation/download, Fecha de consulta: 19/02/2025

[8] Skabar, et al. (2007). Using Argumentation for the descomposition and classification of task for hybrid system development, Progress in Connectionist Based Information Systems. Proceedings of the 2007 International Conference on Neural Information Proccesing and Intelligent Information Systems, Singapur.

[9] Cáceres Nieto, E. (2008). EXPERTIUS: A Mexican Judicial Decision-Support System in the field of family law. In E. Francesconi, G. Sartor, y D. Tiscornia (eds.), Legal Knowledge and Information Systems (pp. 78-87). Frontiers in Artificial Intelligence and Application.

[10] Díaz Dávila, L., Corvalán, juan, Papini, C., Morilla, C., Nieto, F., & Sosa, M. (2022). Arquitectura de Inteligencia Artificial para detectar patrones en documentos escaneados. Simposio Argentino de Informática y Derecho (SID 2022) - JAIIO 51, pag 20-34.

[11] Calliconde Carrasco, Carmela (2022). Detector de mentiras mediante el análisis de estrés de voz usando redes neuronales recurrentes. Universidad Mayor De San Andrés. Facultad De Ciencias Puras Y Naturales. Carrera De Informática. Tesis De Grado.

[12] Diorio, G. (2019) Expedientes electrónicos en las Universidades Nacionales. Info SIU Boletín Electrónico del SIU nro 110. Abril de 2019. Disponible en https://www.siu.edu.ar/rd/InfoSIU/infosiu_110.pdf, Fecha de consulta: 19/02/2025

[13] Sposito O, Bossero J., Ledesma V., Quevedo S. & Matteo L. (2024) Evaluación de un Método Mejorado del Algoritmo K-Means Aplicado en un Sistema de Recuperación de Documentos. CACIC 2024, 570-580. ISBN 978-950-34-2428-5.

Propuesta de validación de datos abiertos y detección de patrones en salud con Machine Learning: Aplicación en el Proyecto INCUCAI (Donación y Trasplante de Órganos y Tejidos)

Roxana Martínez, Santiago Acha, Carlos Bateca, Victoria Agnelli, Eugenio Clavijo, Maximiliano Gribaudo, Agustín Lujan, Sebastián Tsiros, Pedro Negrotto

CAETI - Centro de Altos Estudios en Tecnología Informática
Universidad Abierta Interamericana (UAI)
Montes de Oca 745, Ciudad Autónoma de Buenos Aires, Argentina

{roxana.martinez, santiago.acha}@uai.edu.ar

{CarlosAlberto.BatecaCaicedo, Victoria.Agnelli, Eugenio.Clavijo, MaximilianoWalter.Gribaudo, AgustinEzequiel.LujanBidondo, SebastianMartinDemetrio.Tsiros, PedroGonzalo.Negrotto}@alumnos.uai.edu.ar

RESUMEN

En el contexto de los datos abiertos, la calidad de la información es crucial, especialmente en sectores sensibles como la salud. Aunque existen diversas guías sobre la publicación de datos abiertos, poco se ha explorado sobre la validación específica de los datos en áreas críticas, como el sistema de donación y trasplante de órganos. Este proyecto, en esta línea de investigación, propone un enfoque innovador para la validación de datos abiertos y la detección de patrones mediante técnicas de Machine Learning, aplicados específicamente al Proyecto INCUCAI. Se busca mejorar la integridad y precisión de los datos relacionados con la donación y trasplante de órganos y tejidos, facilitando una toma de decisiones más eficiente y fundamentada. A través de este estudio, se desarrollan métodos para validar conjuntos de datos públicos en el ámbito de la salud y se aplican algoritmos predictivos para identificar patrones relevantes que optimicen los procesos de asignación y distribución de órganos. Esta propuesta no solo contribuirá a la mejora de políticas de salud pública, sino que también apoyará la transparencia en la gestión de datos sensibles. Este trabajo, no solo desarrolla prototipos de software en validaciones de datos específicos, sino que, además, permite estudiar modelos predictivos (aprendizaje automático) que pueden brindar un mejor análisis para la toma de decisiones.

Palabras clave: Datos Abiertos, Métricas de Calidad de Datos, Machine Learning, Predicciones a partir de Patrones.

CONTEXTO

El presente trabajo se enfoca en el proyecto denominado "*Investigación y desarrollo de software para la validación de la calidad de datos abiertos e identificación de patrones para predicciones*", que tuvo inicio en el mes de marzo 2022 y con fecha de culminación en el mes de marzo 2025, y se realizó una extensión de un año más (línea de extensión INCUCAI). Este proyecto pertenece a la línea de investigación de Ingeniería de Software del Centro de Altos Estudios en Tecnología Informática (CAETI) de la Facultad de Tecnología Informática de la Universidad Abierta Interamericana (UAI), el cual contribuye al desarrollo de las Tecnologías de la Información y las Comunicaciones (TIC) en Argentina y en el mundo, llevando adelante la investigación en diversas áreas a través de tratamiento y manipulación de datos en diversas temáticas, tales como: aspectos gubernamentales, gobierno de datos, economía, salud, tecnología, entre otras. El

proyecto es financiado y evaluado por la Secretaría de Investigación de la Universidad, tiene una duración de 4 años, y cuenta con la participación de docentes y estudiantes de grado y posgrado en diversas carreras de la Facultad de Tecnología Informática.

1. INTRODUCCIÓN

Este proyecto abarca varias temáticas con enfoques sobre propuestas, pero en esta oportunidad, se presenta el estudio de la disponibilidad de datos sobre Órganos y Tejidos en la República Argentina. Este análisis se realiza a través del sitio web del Instituto Nacional Central Único Coordinador de Ablación e Implante, INCUCAI [1], y como eje central se espera analizar los datos históricos extraídos del sitio web público del INCUCAI, que permitan resaltar y descubrir tendencias que aporten información y concientización a la población, sobre este contexto y necesidad frente a las largas listas de espera de donantes.

Con este estudio se espera que se identifican algunos factores de ayuda a las personas con base a determinadas situaciones ciudadanas de las distintas provincias de la República Argentina, mediante los datos disponibilizados del sistema de SINTRA [2] que es el Sistema Nacional de Información de Procuración y Trasplante de la República Argentina.

La salud pública es una de las necesidades más esenciales para la humanidad, y los avances en medicina, como los trasplantes de órganos, representan un progreso significativo [3]. Sin embargo, a pesar de los avances tecnológicos, persiste una gran desproporción entre la demanda de trasplantes y la oferta de órganos disponibles, lo que genera largas listas de espera. Este desequilibrio subraya la importancia de mejorar los procesos de selección de donantes, basados en la compatibilidad genética entre donante y receptor, y el papel clave de la comunidad hospitalaria en la detección de posibles donantes [4]. El uso de datos públicos abiertos, combinado con tecnologías emergentes, ofrece una oportunidad para mejorar la calidad y seguridad de los procedimientos de trasplante, permitiendo identificar áreas críticas en la asignación de órganos. Al tratarse de información accesible, los datos abiertos facilitan la interoperabilidad y el análisis para detectar carencias y necesidades específicas en distintos contextos, como la distribución de recursos en zonas geográficas. Este enfoque podría contribuir significativamente a la optimización de los trasplantes y el valor social que estos generan.

Por lo explicado anteriormente, es preciso conocer que "la cultura de datos abiertos trata de obtener un valor añadido de la información […], la información no genera sólo su valor por estar reservada a unos pocos, sino que lo hace por su disponibilidad para ser interpretada y traducida por cualquier actor interesado en trabajar con ella" [5], es decir, un dato abierto puede ser accedido a través de un formato especial que permite la interoperabilidad con otros softwares, y así lograr conocer que tipos de aspectos pueden ser mejorados en la sociedad, por ejemplo, si son necesarios determinados cuidados y/o servicios en provincias específicas debido a la falta de recursos. Todo esto se podría lograr con un estudio de los datos para conocer que falencias y que mejoras son las que se podrían trabajar para lograr un valor agregado a la comunidad.

Existen varios estudios centrados en la medición de la calidad de los datos públicos en estos aspectos [6] [7] [8], mientras que otras investigaciones se enfocan en las métricas utilizadas para evaluar la calidad de estos datos, así como en informes sobre el estado de estos elaborados por organizaciones dedicadas al análisis del impacto de los datos públicos [9].

Este trabajo analizó las expresiones de voluntad registradas por INCUCAI para identificar motivos de desinterés en la población y mejorar la tasa de detección de potenciales donantes. Se constató que disponer de la información del donante con antelación optimiza el proceso de donación, reduciendo la pérdida de órganos. Aumentar la participación ciudadana podría permitir el uso de modelos predictivos para fortalecer la

procuración de órganos. Aunque aún existen grandes desafíos, el estudio muestra que la escasez de órganos se debe principalmente a su rápida caducidad y a que muchos no llegan a ser trasplantados debido a múltiples factores. Mejorar la información disponible y agilizar los trámites médicos podría optimizar este proceso, especialmente con el uso de herramientas adecuadas para el procesamiento de datos. Además, al aumentar la participación de las personas y por ende la cantidad de datos, se podría pensar en modelos predictivos que fortalezcan la actual procuración de órganos.

Si bien como ya se mencionó, aún falta un largo camino, en este trabajo se pudo analizar que lograr identificar más información en aspectos a futuros, agilizaría los tramites médicos y potenciaría mucho más la procuración actual de órganos, en conjunto con la utilización de las herramientas adecuadas para el procesamiento de datos.

2. LÍNEAS DE INVESTIGACIÓN DESARROLLO

Este proyecto pertenece a la línea de investigación de Ingeniería de Software del Centro de Altos Estudios en Tecnología Informática (CAETI). Para esta última instancia del proyecto, se hace enfoque en los siguientes ejes:

- Desarrollo de algoritmos de validación y detección de patrones en datos públicos abiertos en el ámbito de la salud, con especial atención al Proyecto INCUCAI y su relación con la donación y trasplante de órganos y tejidos.
- Culminar con el diseño y desarrollo de algoritmos para identificar patrones en datos públicos abiertos orientados en temáticas de salud, políticas de buenas prácticas en calidad de datos, tecnologías, y aspectos internacionales que puedan llevar a cabo un valor agregado a la comunidad.
- Analizar los modelos predictivos orientados a contextos gubernamentales, como así también, analizar posibles predicciones sobre nuevos datos encontrados con técnicas de machine learning.
- Desarrollo de nuevas funcionalidades para herramientas de validación de datasets públicos, basadas en métricas de calidad de datos específicas para el sector salud, como la consistencia y la integridad de los datos geoespaciales.
- Análisis de casos predictivos y aplicación con casos reales y prototipos desarrollados.
- Elaboración de una guía de buenas prácticas para el uso de Machine Learning en la validación y análisis de datos abiertos en el ámbito de la salud.

3. RESULTADOS OBTENIDOS/ESPERADOS

Enfoque del proyecto:

En el contexto del Gobierno Abierto, el acceso y análisis de datos públicos en áreas críticas como la salud pública juegan un papel fundamental en la toma de decisiones basada en evidencia. El Proyecto de investigación, presenta en una de sus líneas de trabajo, una iniciativa basada en aspectos de análisis de daos públicos del INCUCAI [1], el cual gestiona información sobre la donación y trasplante de órganos y tejidos en la República Argentina. Esta propuesta, nace de la necesidad de requerir estrategias que permitan validar la calidad de los datos públicos que brinda en su sitio web [2] y mejorar su aprovechamiento mediante técnicas de predicción basadas en Machine Learning.

Este trabajo se centra en la optimización del proceso de recolección, validación y análisis de datos provenientes del sitio web del INCUCAI [2], con énfasis en la aplicación de modelos predictivos que permitan identificar tendencias y patrones en la disponibilidad de órganos y tejidos. Se propone la implementación de técnicas de validación para garantizar la integridad de la información extraída, así como el desarrollo de enfoques analíticos que optimicen la generación de predicciones sobre la evolución de la donación y trasplante en Argentina.

Objetivos principales:
Este trabajo propone un análisis sobre los alcances, desafíos y mejores prácticas a partir de casos prácticos y datos reales con los que se trabajan en Gobierno Abierto. El objetivo principal para esta línea de investigación del Proyecto es analizar y mejorar la calidad de los datos abiertos relacionados con la donación de órganos y tejidos, proponiendo estrategias de validación y herramientas de análisis predictivo que permitan extraer información confiable para estudios en salud pública.

Específicamente, se busca:

- Desarrollar técnicas de validación para garantizar la calidad e integridad de los datos extraídos del sitio web del INCUCAI.
- Comparar diferentes enfoques de recolección y estructuración de datos mediante herramientas de web scraping y APIs desarrolladas en Python y C#.
- Implementar modelos de predicción con Machine Learning para analizar tendencias en la disponibilidad de órganos y tejidos.
- Evaluar la efectividad de diferentes herramientas analíticas, como Power BI, Excel y Python, en la exploración y visualización de patrones de datos en salud pública.

Metodología y Técnicas:
El proceso metodológico de investigación empleado en esta etapa del proyecto se basa en un enfoque sistemático cualitativo y cuantitativo, estructurado en varias fases que permiten la validación de datos públicos y la identificación de patrones mediante técnicas de Machine Learning. Inicialmente, se lleva a cabo la extracción automatizada de datos del sitio web del INCUCAI mediante técnicas de web scraping y el desarrollo de APIs en Python y C#. Posteriormente, se realiza un proceso de validación para garantizar la calidad e integridad de los datos recolectados, asegurando su consistencia y disponibilidad para análisis posteriores. Luego, estudio de escenarios de identificación de patrones sobre éstos con enfoques predictivos sobre nuevos datos; desarrollo de prototipos (análisis cuantitativo de datos); visualización de la predicción de patrones; validación de la solución propuesta; identificación de las limitaciones del trabajo.

Cabe mencionar que los datos son analizados y estructurados, aplicando técnicas de limpieza y transformación para facilitar su procesamiento en modelos predictivos. Y sobre eso, se efectúa un estudio detallado de los escenarios posibles para la identificación de patrones en la evolución de la donación y trasplante de órganos y tejidos, considerando variables como el tiempo, la ubicación geográfica y las tendencias en la demanda de trasplantes.

Resultados obtenidos:
En las primeras etapas del proyecto, se llevó a cabo un análisis exhaustivo de las deficiencias en la calidad de los datos públicos abiertos de diversos sitios web, identificando técnicas de validación adecuadas para garantizar su integridad en formatos estándar. Además, se realizó un estudio detallado de los datos públicos geolocalizados [10] con el objetivo de optimizar su uso en distintas aplicaciones. Si bien el proyecto abarca diversas áreas de datos, como tecnología, salud y financiamiento, en esta fase se ha centrado específicamente en el ámbito de la salud pública, con un enfoque en el estudio de la disponibilidad de órganos y tejidos en Argentina a través del análisis del sitio web del INCUCAI.

Algunos aspectos de resultados obtenidos en esta línea de INCUCAI, son las siguientes: en las primeras etapas del estudio, se han desarrollado y comparado distintas herramientas para la extracción automatizada de datos desde el sitio web del INCUCAI, asegurando la integridad y disponibilidad de la información para su posterior análisis [11]. Además, se ha implementado web scraping utilizando Python y C#, evaluando el rendimiento y la eficiencia de cada enfoque en términos de tiempos de ejecución, consumo de recursos y facilidad de

implementación [12]. Otro aspecto analizado, fue que se han explorado alternativas de bajo costo para la obtención de datos públicos, demostrando la viabilidad de metodologías accesibles para organizaciones con recursos limitado [13].

Es importante mencionar que actualmente se están evaluando distintos modelos predictivos mediante herramientas de software y teniendo en cuenta varios aspectos sobre este contexto [14], con el fin de optimizar la generación de predicciones y mejorar la precisión del análisis de datos, aplicando técnicas de Machine Learning para la obtención de resultados más precisos y confiables.

4. FORMACIÓN DE RECURSOS HUMANOS

Este proyecto se compone por una docente investigadora, Doctora en Ciencias Informáticas de la Universidad Nacional de La Plata (UNLP), directora del Proyecto de Investigación, 1 (un) docente auxiliar de la carrera Ingeniería en Sistemas Informáticos y 2 (dos) estudiantes tesistas de la carrera de Maestría en Tecnología Informática en la Universidad Abierta Interamericana (UAI). También cuenta con un docente tesista de la carrera del Doctorado en Ciencias Informáticas de UAI, 4 (cuatro) estudiantes colaboradores de la carrera de Ingeniería en Sistemas Informáticos de UAI, y 1 (un) tesistas de la carrera Licenciatura en Gestión de Tecnología Informática en UAI. En relación directa con la línea de I/D presentada para el proyecto, los miembros del equipo se encuentran en realización de: 1 tesis doctoral, 2 tesis de maestría y 1 tesinas de grado de la UAI.

5. BIBLIOGRAFÍA

[1] Argentina.gob.ar - INCUCAI (2025). "*INCUCAI*". Disponible en: https://www.argentina.gob.ar/salud/incucai

[2] INCUCAI - SINTRA (2025). "*SINTRA - El Sistema Nacional de Información de Procuración y Trasplante de la República Argentina*". Disponible en: https://sintra.incucai.gov.ar/

[3] Méndez-Martínez, F. (2021). Los Retos de la Donación de Órganos Durante el Brote de COVID-19. Revista Salud y Administración, 8(24), 45-51.

[4] Defelitto, J. (2011). Trasplante de órganos: generalidades. PARTE III, 540. Disponible en: https://acortar.link/wfPyQB

[5] Magallón Rosa, R. (2017). Datos abiertos y acceso a la información pública en la reconstrucción de la historia digital.

[6] ISO 25012 (2008). Ingeniería de software - Requisitos de calidad y evaluación de productos de software (SQuaRE) - Modelo de calidad de datos.

[7] Martínez, R. et al. (2021). Metrics proposal to measure the quality of governmental datasets. IEEE Latin America Transactions, Vol. 100. ISSN 1548-0992.

[8] de España, G. (2017). Manual práctico para mejorar la calidad de los datos abiertos. Madrid.

[9] ODI Open Data Institute. "*The 2019 Data Skills Framework*". Disponible en: https://theodi.org/article/open-data-skills-framework/

[10] Martínez, R., & Simón, A. (2023). Analysis of the Types of Argentine Geospatial Public Open Data. *Engineering Proceedings*, 42(1), 7.

[11] Agnelli, V., Martínez, R. et al. (2024). *Análisis Comparativo del Rendimiento de C# y Python para Web Scraping: El Caso de la Extracción de Datos de INCUCAI.* Congreso Nacional de Ingeniería Informática – Sistemas de Información (CoNaIISI). Universidad Nacional de Catamarca. Argentina.

[12] Bateca, C., Martínez, R. et al. (2024). *Evaluación de Herramientas para Predicciones en Regresión Lineal: Estudio con Datos del INCUCAI*. Congreso Argentino de Ciencias de la Computación (CACIC).

[13] Martínez, R., Acha, S. et al. (2024). *Webscrapping lowcost. Caso INCUCAI*. Congreso Internacional en Innovación Tecnológica Informática. CIITI.

[14] Cyunel, M., Garea, M., & Neira, P. (2019). Procuración y trasplante de órganos y tejidos. Rev Hosp Niños (B Aires), 61(273), 88-94.

Tendencias Actuales sobre Bases de Datos no Convencionales

M. D. Alba, J. Arroyuelo, M. E. Di Genaro, A. Grosso, M. Jofré, V. Ludueña, N. Reyes, D. Welch
Dpto. de Informática, Fac. de Cs. Físico-Matemáticas y Naturales, Universidad Nacional de San Luis
{*bjarroyu, digeme, agrosso, vlud, nreyes,dwelch*}@unsl.edu.ar,{*mdaniela.alba, monicajofre*}@gmail.com
Edgar Chávez
Centro de Investigación Científica y de Educación Superior de Ensenada, México
elchavez@cicese.mx
Karina Figueroa
Fac. de Cs. Físico-Matemáticas, Universidad Michoacana de San Nicolás de Hidalgo, México
karina@fismat.umich.mx
Rodrigo Paredes
Fac. de Ingeniería, Universidad Finis Terrae, Chile
rparedes@uft.cl

Resumen

El avance tecnológico ha influido en todos los aspectos de la vida diaria, generando una gran cantidad de datos digitales provenientes de diversas fuentes, cada una con sus características propias. Éstos incluyen desde datos obtenidos a través de satélites, hasta información de sensores variados, investigaciones médicas, secuencias biológicas, entre otros. Debido a su complejidad, estos datos requieren un procesamiento especializado para satisfacer las diversas necesidades del mundo contemporáneo. La adopción de bases de datos que emplean *espacios métricos* para gestionar estos objetos complejos y no estructurados, surge como una respuesta directa a la variedad y volumen de los mismos.

El desarrollo en pos de la consolidación de este modelo de bases de datos es el objetivo de nuestro proyecto. Así, se trabaja en sistemas que no sólo resuelvan sus limitaciones actuales, sino que también prevean las necesidades futuras, asegurando que sean escalables, seguros y capaces de gestionar los posibles datos emergentes. En resumen, la evolución de las *Bases de Datos Métricas* es un paso crucial para adaptarse al rápido crecimiento de datos digitales, optimizando no solo la capacidad de almacenamiento y recuperación de información, sino abriendo también nuevas oportunidades para analizar y generar conocimiento a partir de esos datos.

Palabras Claves: bases de datos métricas, índices, búsquedas por similitud.

Contexto

Este trabajo tiene como objetivo abordar diversos aspectos de las Bases de Datos Métricas, consideradas como un modelo de Bases de Datos no Convencionales, que aún presentan desafíos sin resolver. En este contexto, se desarrollan nuevos índices orientados a memorias jerárquicas, diseñados para ser dinámicos, escalables y eficientes en las operaciones de E/S. Además, se llevan a cabo estudios sobre arquitecturas de procesadores que puedan beneficiar y optimizar este modelo.

Los estudios y desarrollos detallados en este artículo se enmarcan en el Proyecto Consolidado *Tecnologías Avanzadas de Bases de Datos*, en particular la línea *Bases de Datos no Convencionales*. El mismo se ejecuta en la Universidad Nacional de San Luis (PROICO 03-1723), pertenece al Programa de Incentivos y está incorporado al Laboratorio de Investigación y Desarrollo en Bases de Datos (LaBDa), en el que colaboran investigadores de otros grupos de la región: Universidad Finis Terrae (Chile), Universidad Michoacana de San Nicolás de Hidalgo y Centro de Investigación Científica y de Educación Superior de Ensenada (México).

Introducción

El uso masivo de dispositivos capaces de producir datos digitales ha provocado un crecimiento sostenido tanto en la magnitud como en la diversidad de los mismos. Originados en sectores tan variados como el industrial, laboral, sanitario, educativo, científico, etc., su número ha crecido de manera exponencial.

Este panorama vuelve crucial la evolución de las bases de datos especializadas en el manejo de *datos no estructurados.* Éstas se han tenido que adaptar

rápidamente para gestionar de manera efectiva tanto el volumen como la heterogeneidad de los datos. Por otro lado, las aplicaciones modernas requieren operaciones particulares sobre estos datos, tales como encontrar las huellas digitales más similares a una dada o la identificación de patrones digitales similares. En estos escenarios, las búsquedas tradicionales basadas en coincidencias exactas resultan insuficientes, siendo las *búsquedas por similitud* mucho más adecuadas. Estas búsquedas se basan en un objeto de referencia para localizar otros elementos "similares" dentro de la base de datos, y se conocen como consultas por contenido o consultas mediante ejemplo (query by example).

Una metodología altamente eficaz para aplicaciones que demandan búsquedas por similitud es el *modelo de espacios métricos*. Este enfoque, aplicable a una amplia variedad de entornos, se basa en un conjunto de objetos y una función de distancia que evalúa las disimilitudes entre ellos. El modelo puede emplearse para cualquier tipo de objeto no estructurado que pueda compararse mediante una medida de distancia, siempre que dicha medida cumpla con las propiedades de una métrica. Generalmente, estas métricas son definidas por especialistas en el tipo de datos que se quiere comparar.

Para gestionar de manera eficiente las diversas demandas asociadas a los objetos almacenados en estas bases de datos, es necesario emplear los *Métodos de Acceso Métricos* (MAMs), los cuales permiten realizar búsquedas sin revisar toda la base de datos de manera secuencial, mejorando así la eficiencia en la recuperación de información [5, 9]. Dado el constante avance de los sectores que aplican este modelo, es esencial actualizar y optimizar continuamente los MAMs, adaptándolos a cada situación particular y enfrentando desafíos como la gestión de grandes volúmenes de datos, la resolución de consultas cada vez más complejas y las actualizaciones (inserciones/eliminaciones).

Líneas de Investigación y Desarrollo

Bases de Datos no Convencionales

Como se expresó anteriormente, las bases de datos capaces de administrar vídeos, texto libre, imágenes, secuencias de proteínas o ADN, audio, etc., son formalizadas utilizando el *modelo de espacios métricos*, por lo que también se las conoce como *Bases de Datos Métricas*.

En este modelo, calcular la distancia entre dos objetos puede resultar en un alto costo computacional, por lo que el número de cálculos realizados durante alguna operación se usa habitualmente como medida de complejidad. Estos cálculos son fundamentales tanto para la creación de índices como para la realización de consultas sobre ellos. Por lo tanto, para responder de manera efectiva las diversas consultas evitando comparaciones exhaustivas sobre la base de datos, es esencial implementar índices especializados. La optimización de estos índices se ha convertido en un objetivo clave, centrándonos en aquellos que han mostrado eficiencia, principalmente en las búsquedas, para reducir su complejidad. En este proceso se toman en cuenta factores como el nivel de memoria donde se alojarán, su capacidad para adaptarse a modificaciones y su escalabilidad, entre otros.

Formalmente un espacio métrico se define mediante un conjunto $\mathbb{U}$ (Universo) y una función de distancia $d : \mathbb{U} \times \mathbb{U} \rightarrow \mathbb{R}^+$ que satisface condiciones como: reflexividad, positividad estricta, simetría y desigualdad triangular, lo que resulta muy útil con la resolución de consultas por similitud. Dos de esas propiedades son esenciales para el ahorro de cálculos de distancia: la simetría y la desigualdad triangular. Luego, una base de datos X es un subconjunto de $\mathbb{U}$ y las consultas por similitud se realizan a partir de un elemento $q \in U$, pudiendo ser consultas *por rango* (q, r) o de *k-vecinos más cercanos* k-NN(q)).

Los k Vecinos más Cercanos

En el contexto de consultas por similitud en espacios métricos, la búsqueda de los k-vecinos más cercanos (k-NN) es sumamente valiosa; un ejemplo práctico es la localización de hoteles cercanos a una ubicación específica. Una variante importante de esta búsqueda es la consulta *All-k-NN*, que consiste en asociar cada elemento $u \in X$ con los k elementos más cercanos a él dentro del conjunto X-$\{u\}$. La solución más sencilla, que compara cada objeto con todos los demás, tiene una complejidad de $O(n^2)$ (con $|X|$=n), lo que resulta ineficiente para conjuntos de datos grandes. Para optimizar este proceso, se procede a preprocesar la base de datos construyendo un índice para luego buscar en el mismo los k-NN de cada elemento del conjunto, permitiendo reducir el número de comparaciones necesarias y mejorar el rendimiento de las consultas.

Sin embargo, existen situaciones en las cuales el costo de esta solución puede resultar tan excesivo como la solución ingenua. Éste es el caso de gestionar una base de datos masiva, o si la función de distancia es demasiado costosa de calcular, o inclu-

so si se está trabajando con espacios métricos de alta dimensión. Para tales casos, se pueden considerar búsquedas por similitud *aproximadas*, en las cuales se sacrifica precisión por velocidad; es decir, se mejora la complejidad aceptando algunos "errores" en la respuesta [14, 5, 15, 10].

Para abordar estas problemáticas, se han propuesto métodos que calculan los *All-k-NN sin la necesidad de un índice* y que permiten obtener además una aproximación del *Grafo de los k-vecinos más cercanos* (kNNG) [13]. Éstos conectan cada objeto u de la base de datos con k vecinos *cercanos*, relajando la condición que exige que no haya, algún objeto más cercano a u que los k vecinos devueltos. Esto puede provocar que la respuesta pierda algún objeto muy cercano, devolviendo uno más lejano en su lugar, a cambio de una respuesta más rápida. Este grafo se denominó *Grafo de vecinos cercanos* (knNG) [3].

Una de las estrategias propuestas aborda el caso particular de k=1; ésta utiliza la información del proceso de *construcción* del *DiSAT* [4] para generar el 1nNG, conectando cada objeto con un elemento cercano, que puede ser, o no, su vecino más cercano [3]. Este método logra un equilibrio entre calidad y tiempo de respuesta, logrando muy buena precisión con un bajo error, *sin depender de búsquedas en índices*.

También se desarrollaron otras tres estrategias que abordan el problema general de *All-k-NN* y calculan el k-NNG sin depender de índices. Estos algoritmos aprovechan estratégicamente las propiedades de la función de distancia, seleccionan *muestras* de la base de datos y calculan distancias iniciales, para luego usar los principios de simetría y desigualdad triangular para identificar los vecinos exactos [2] o aproximados [8] de todos los objetos en la base de datos.

Índices Métricos

Como se ha señalado, optimizar los índices métricos es uno de los objetivos principales de esta línea de investigación. Dado el amplio rango de aplicaciones de estos índices, es necesario considerar diversas perspectivas para lograr su optimización. Se deben tomar en cuenta aspectos como su escalabilidad, dinamismo, y su capacidad para adaptarse al almacenamiento en memoria secundaria, entre otros factores. En este último caso, es crucial ajustar la medida de complejidad, teniendo en cuenta no solo la cantidad de cálculos de distancia, sino también el número de operaciones de entrada/salida requeridas.

En cuanto a la gestión de índices, un reto común es que éstos pueden superar la capacidad de la memoria principal debido al tamaño de los objetos indexados o la gran cantidad de objetos almacenados. Para enfrentar este desafío, es esencial diseñar Métodos de Acceso Métrico (MAMs) que mantengan su eficiencia, incluso cuando se almacenan en memoria secundaria. Actualmente, se está trabajando en una versión dinámica del *DDiSAT* optimizada para memoria secundaria. Esta versión busca equilibrar la eficiencia de las búsquedas y los costos de reconstrucción con un uso eficiente del espacio en disco y la minimización de operaciones de entrada/salida (E/S). El objetivo es alcanzar un rendimiento en los cálculos de distancia comparable al de la versión en memoria principal, asegurando una alta ocupación de las páginas de disco, reduciendo las operaciones de E/S y manteniendo la localidad de acceso.

A pesar de la importancia del dinamismo en un índice, existen pocos dinámicos, siendo uno de los más representativos el *Árbol de Aproximación Espacial Dinámico (DSAT)* [12], que se basa en el *Árbol de Aproximación Espacial (SAT)*, uno de los índices más eficientes en espacios métricos de dimensiones medias a altas. El *DSAT* ofrece un mejor rendimiento en las búsquedas, aunque requiere la configuración de un parámetro adicional. Siguiendo el enfoque del *DSAT*, se desarrolló el *Árbol de Aproximación Espacial Distal (DiSAT)* [4], que, además de no necesitar parámetros extra, mejora el rendimiento de búsqueda con respecto a sus predecesores, el *SAT* y el *DSAT*. Sin embargo, el *DiSAT* sigue siendo estático. Posteriormente, se propuso la *Foresta de Aproximación Espacial Distal (DiSAF)* [1], un índice dinámico diseñado para memoria principal. Este índice aplica la técnica de dinamización de Bentley y Saxe al *DiSAT* y aprovecha el extenso conocimiento sobre aproximación espacial para maximizar su rendimiento. No obstante, sus costos de construcción son elevados, ya que requiere la reconstrucción de subárboles tras cada inserción.

Para abordar este problema, se diseñó el *Árbol de Aproximación Espacial Distal Dinámico (DDiSAT)* [7, 6], el cual emplea la técnica de *inserción perezosa* para amortiguar los costos de reconstrucción y mejora el rendimiento en las búsquedas sin requerir espacio adicional. En lugar de insertar elementos de manera inmediata, se retrasan y se colocan en una bolsa asociada a su nodo más cercano en el árbol. Cuando la cantidad de elementos pendientes de inserción en las bolsas alcanza el número total de nodos del árbol, se realiza una reconstrucción completa

del árbol.

El *DDiSAT* reduce significativamente los costos de construcción en comparación con el *DiSAF* y ofrece un mejor desempeño en las búsquedas que los índices estáticos (*SAT* y *DiSAT*) y el índice dinámico *DSAT*. Esto se debe a la efectiva partición del espacio y la distribución óptima de los elementos realizada por el índice. Se ha demostrado que, con la construcción de un *DiSAT* con pocos elementos que permitan particionar el espacio, y agregando el resto de los elementos en las bolsas, se puede lograr un excelente rendimiento en las búsquedas mientras se reducen los costos de construcción. Se ha diseñado un algoritmo de carga masiva para la creación del *DDiSAT*, que minimiza algunas de las reconstrucciones y se combina con la inserción diferida y un algoritmo eficiente para resolver consultas k-NN.

Además, se están ampliando las capacidades de PostgreSQL al integrar algunos de estos índices métricos y sus operaciones, adaptándolo para gestionar y procesar grandes bases de datos que contienen tipos de datos métricos.

Ciencia de Datos Aplicada

La incorporación de la tecnología en la agricultura ha dado lugar a un nuevo paradigma que optimiza la producción a través de herramientas de Ciencia de Datos. En este contexto, la investigación se centra en desarrollar un marco sólido que permita utilizar modelos analíticos y predictivos para mejorar la optimización, eficiencia y sostenibilidad en la producción agrícola. La integración de modelos avanzados facilita el análisis de grandes conjuntos de datos, proporcionando información valiosa (insights) que impulsa decisiones estratégicas, impactando positivamente la productividad, la gestión de recursos y la resiliencia ante cambios climáticos.

La aplicación de Ciencia de Datos en la agricultura abarca diversas áreas, como la predicción de rendimientos y la gestión eficiente de insumos. Esto incluye el análisis de la variabilidad y características de los suelos, el clima y el uso del agua para riego, un recurso costoso que requiere administración cuidadosa. Además, se enfoca en la identificación temprana de enfermedades y plagas que podrían afectar significativamente los cultivos. Este contexto no solo busca aumentar la eficiencia en la producción y el uso de recursos, sino también promover prácticas agrícolas sostenibles, alineándose con los objetivos de desarrollo sostenible establecidos por la Organización de las Naciones Unidas [1].

Investigando Arquitecturas de Procesadores

Una experiencia muy interesante, orientada a los estudiantes, surgió mientras se estudiaba cómo aprovechar las arquitecturas de procesadores para mejorar el rendimiento de las bases de datos. Se propuso una arquitectura evolutiva que comienza con una simple basada en la *MU0* (**M**anchester **U**niversity versión **0**), la cual permite enfrentar desafíos de implementación de estructuras de datos. Progresivamente, se introducen extensiones que facilitan el manejo de vectores y pilas, y se avanza hacia modos de direccionamiento más complejos que permiten que el código sea re-entrante, una característica clave para sistemas operativos. La arquitectura final permite diferenciar entre registros de datos y de dirección, evolucionando hacia un diseño de registros generales y terminando en una arquitectura LOAD/STORE utilizada por la plataforma EDU-CIAA basada en ARM Cortex-M4. Para apoyar el aprendizaje de los estudiantes, se han desarrollado ensambladores y simuladores con interfaces similares a *gdb* de Linux, utilizando herramientas como *lex* y *bison*. La experiencia se extiende al desarrollo de implementaciones en FPGA, permite a los estudiantes interactuar con hardware y programar rutinas de manejo de interrupciones.

Resultados y Objetivos

Mediante el estudio sobre el modelo de espacios métricos, se ha logrado optimizar el rendimiento de los MAMs analizados, abriendo caminos para su implementación en otros métodos de acceso [7, 6, 8, 2, 3, 11, 1].

Se está trabajando en brindar nuevas herramientas que permitan una administración para bases de datos métricas mucho más eficiente, acercándolas a la de los modelos tradicionales de base de datos. Para ello, se profundiza el estudio de nuevos diseños de estructuras de datos enfocados en incrementar su eficiencia en espacio/tiempo, analizando cómo se adaptan al nivel de la jerarquía de memorias en la que se almacenan y a las características de los datos a ser indexados. Al mismo tiempo, se continúa experimentado con técnicas que no utilicen índices pero permitan resolver consultas de manera eficiente. En otro contexto se espera seguir investigando sobre distintos tipos de arquitecturas de procesadores.

[1] https://www.un.org/sustainabledevelopment/es/

Actividades de Formación

Dentro de esta línea de investigación se forman alumnos y docentes-investigadores participando en:

- **Maestría en Cs. de la Computación** (UNSL): tesis sobre una versión dinámica eficiente del *DiSAT*.
- **Maestría en Ciencia de Datos e Innovación Empresarial** (CAECE): tesis sobre la aplicación de la Ciencia de Datos a la producción agrícola.
- **Maestría en Informática** (UNSJ): tesis sobre la evaluación del knNG para búsquedas por similitud.
- **Maestría en Informática** (UNSJ): tesis sobre una versión dinámica y para memoria secundaria del *DDiSAT*.
- **Maestría en Ing. de Software** (UNSL): tesis sobre la incorporación de índices métricos a un DBMS.

Referencias

[1] E. Chávez, M. Di Genaro, N. Reyes, and P. Roggero. Decomposability of disat for index dynamization. *Computer Science & Technology*, pages 110–116, 2017.

[2] E. Chávez, V. Ludueña, and N. Reyes. Solving all-k-nearest neighbor problem without an index. In *Procs. del XXV Congreso Argentino de Ciencias de la Computación CACIC 2019)*, pages 567–576. UniRío editora, 2019.

[3] E. Chávez, V. Ludueña, N. Reyes, and F. Kasián. All near neighbor graph without searching. *Computer Science & Technology*, 18:61–67, 2018.

[4] E. Chávez, V. Ludueña, N. Reyes, and P. Roggero. Faster proximity searching with the distal {SAT}. *Information Systems*, 59:15 – 47, 2016.

[5] E. Chávez, G. Navarro, R. Baeza-Yates, and J. Marroquín. Searching in metric spaces. *ACM Computing Surveys*, 33(3):273–321, September 2001.

[6] Edgar Chávez, María E. Di Genaro, and Nora Reyes. An efficient dynamic version of the distal spatial approximation trees. In *Actas del XXVIII Congreso Argentino de Ciencias de la Computación (CACIC 2022)*, pages 468–477, Oct. 2022.

[7] Edgar Chávez, María E. Di Genaro, and Nora Reyes. Dynamic distal spatial approximation trees. In Patricia Pesado, editor, *Computer Science – CACIC 2022*, pages 175–189, Cham, 2023. Springer Nature Switzerland.

[8] Edgar Chávez, Verónica Ludueña, and Nora Reyes. Heuristics for computing k-nearest neighbors graphs. In Patricia Pesado and Marcelo Arroyo, editors, *Computer Science – CACIC 2019*, pages 234–249, Cham, 2020. Springer International Publishing.

[9] Lu Chen, Yunjun Gao, Xuan Song, Zheng Li, Yifan Zhu, Xiaoye Miao, and Christian S. Jensen. Indexing metric spaces for exact similarity search. *ACM Comput. Surv.*, 55(6), dec 2022.

[10] P. Ciaccia and M. Patella. Approximate and probabilistic methods. *SIGSPATIAL Special*, 2(2):16–19, 2010.

[11] K. Figueroa, N. Reyes, A. Camarena-Ibarrola, and L. Valero-Elizondo. Improving the list of clustered permutation on metric spaces for similarity searching on secondary memory. In *10th Mexican Conference on Pattern Recognition (MCPR2018)*, volume 10880, pages 82–92, 2018.

[12] G. Navarro and N. Reyes. New dynamic metric indices for secondary memory. *Information Systems*, 59:48 – 78, 2016.

[13] R. Paredes, E. Chávez, K. Figueroa, and G. Navarro. Practical construction of k-nearest neighbor graphs in metric spaces. In *Proc. 5th Workshop on Efficient and Experimental Algorithms (WEA)*, LNCS 4007, pages 85–97, 2006.

[14] H. Samet. *Foundations of Multidimensional and Metric Data Structures (The Morgan Kaufmann Series in Computer Graphics and Geometric Modeling)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2006.

[15] P. Zezula, G. Amato, V. Dohnal, and M. Batko. *Similarity Search: The Metric Space Approach (Advances in Database Systems)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005. XVIII, 220 p., Hardcover ISBN: 0-387-29146-6.

Visualizando los datos de DiGiBePe

Prog. Luis Olguin, Mag. Raúl Klenzi
Instituto de Informática – Departamento de Informática
Facultad de Ciencias Exactas, Físicas y Naturales
Universidad Nacional de San Juan
lolguin@iinfo.unsj.edu.ar; rauloscarklenzi@unsj-cuim.edu.ar

Resumen

DiGiBePe es el nombre adoptado por la Comisión Nacional de Bibliotecas Populares (CONABIP) para el Sistema Integrado de Bibliotecas KOHA. Este software libre permitió a las bibliotecas tener presencia en la web y desde allí administrar sus acervos bibliográficos, estadísticas, socios, etc.

La Sociedad Franklin Biblioteca Popular es la biblioteca pionera en la provincia de San Juan en adoptar este software, pero luego de más de 10 años de uso nunca ha encarado un proyecto de análisis de los datos, que en su sistema se almacenan, para la toma de decisiones.

El proyecto **Datos Populares** surgió con el objetivo de colaborar con la biblioteca Franklin a fin de formar recursos humanos que sean capaces de interactuar con herramientas de **visualización de datos** con el objeto de aplicarlas luego para obtener vistas gráficas que permitan analizar distintos aspectos del día a día de una biblioteca popular, tomando como fuente de datos las tablas Mysql que posee el sistema **DiGiBePe instalado en el dominio de la Biblioteca Popular Franklin de San Juan (http://0704.bepe.ar)**.

Este artículo detalla los logros obtenidos hasta el momento y las perspectivas a futuro para la mejora en la toma de decisiones bibliotecarias basadas en datos.

Contexto

Este trabajo se enmarca en las actividades establecidas en el proyecto "Datos Populares" (PDTS 21/E1246), ejecutado por el Instituto de Informática y el Departamento de Informática de la Facultad de Ciencias Exactas, Físicas y Naturales de la Universidad Nacional de San Juan, aprobado con evaluación externa y contenido en el marco del Laboratorio de Bibliotecas Digitales y el Laboratorio de sistemas inteligentes para la búsqueda de conocimiento en datos masivos del Instituto de Informática.

1. Introducción

Visualizar datos extraídos desde la base de datos de DiGiBePe permite a una biblioteca popular la comprensión de las características de su catálogo, descubrir patrones de uso de las colecciones por parte de sus usuarios y lo más importante, abre el diálogo entre los bibliotecarios para la toma de decisiones, por ejemplo, acerca de la posibilidad de expurgo en determinadas áreas con alta densidad de obras, pero con poca demanda en la actualidad.

(OCLC Research, 2022) realizó una encuesta a bibliotecas donde surge que la mayoría de las instituciones participantes consideran que están *inundadas de datos* lo que genera la sensación de que se trata de elementos difíciles de manejar, limpiar y usar eficientemente. Destacan que es fundamental obtener información a

partir de los datos para así optimizar el uso de sus presupuestos, por ejemplo, para adquisiciones.

(Phetteplace 2016) destaca que una representación gráfica permite al cerebro humano asimilar con mayor rapidez la información que se intenta difundir, comparado con el trabajo de selección e identificación a partir de información solo textual o numérica. La posibilidad de generar gráficas con los datos obtenidos revela relaciones que de otra manera son difíciles de descubrir. Identifica los pasos que se deberían cumplimentar para alcanzar una visualización exitosa, entre los que se encuentran los que se han llevado adelante en el *proyecto PDTS Datos Populares*: adquirir, analizar, filtrar, extraer, representar, refinar e interactuar.

(FESABID, 2014) generó un estudio de usuarios que de alguna manera coincide que con los resultados de las encuetas realizadas en el marco del proyecto PDTS Datos Populares, donde se observa que el usuario de una biblioteca popular valora la posibilidad de asistencia profesional bibliotecaria, comodidad y acceso a los espacios de lectura, actividades culturales, en definitiva, que la biblioteca facilite el acceso a la información y a la cultura de manera libre y gratuita.

El grupo de trabajo del Laboratorio de Grandes Datos analizó distintas herramientas de visualización de datos (Olguin, 2016), tal es el caso de NodeXL, Ghephi y Tableau. Este último software, en su versión gratuita, fue el seleccionado para realizar las capacitaciones en visualización de datos, pero dada la dificultad en el aprendizaje de este soft, se optó por iniciar el camino de las visualizaciones utilizando Excel y complementarlas con Tablas Dinámicas.

2. Líneas de Investigación y Desarrollo

El proyecto **PDTS Datos Populares (**21/E1246 – UNSJ 2023) busca *diseñar un panel de visualización de datos para bibliotecas populares* que permita a los integrantes de la biblioteca (directivos y bibliotecarios) mirar los datos almacenados en DiGiBPe y utilizarlos posteriormente para mejorar los servicios que ofrece la biblioteca.

3. Resultados Obtenidos/Esperados

En el periodo Junio/23 – Diciembre/24 el equipo interdisciplinario del proyecto ha logrado avances significativos para *conocer el día a día de la biblioteca*.

Se trabajó junto a los bibliotecarios de la Sociedad Franklin para capacitarlos en la extracción de datos desde las tablas de *Socios, Préstamos, Estadísticas y Catálogo* del sistema DiGiBePe. Resultado de esto, se han conformado un conjunto de *informes KOHA-SQL* que permiten obtener datos claves sobre el uso de la biblioteca, facilitando su transformación en información accionable. Es de destacar que los informes generados hasta la fecha responden a lo requerido por los bibliotecarios para conocer acerca de sus datos.

Destacamos, entre los informes generados, los siguientes:

Tamaño de la colección: Por primera vez los bibliotecarios pudieron analizar cada bloque de su colección y el peso de cada sección en el total. Para este análisis

visual se organizaron los datos por Clasificación Decimal de Dewey.

Circulación de la Colección: Permite visualizar en qué parte de la colección hacen foco los usuarios, analizando el comportamiento de los préstamos. Por primera vez se pudo mostrar cómo han ido cambiando las preferencias lectoras en la biblioteca popular y cuáles son las colecciones con más requerimientos, dato importante al momento de planificar nuevas adquisiciones si tenemos en cuenta los recursos escasos con los que cuentan estas organizaciones.

Antigüedad del Catálogo: Se generó una vista gráfica en donde se observa el porcentaje de obras con edición más reciente. Dato importante al momento de analizar la satisfacción de los usuarios respecto a la sección novedades. Las bibliotecas populares reciben muchas obras en donación que no siempre son de ediciones recientes. Contar con esta vista les permite a los bibliotecarios decidir si la obra recibida puede sumarse tal vez a una colección poco reforzada en el catálogo.

Ciclo de Vida de un Título: Permite generar una vista gráfica de cuál es el comportamiento de los usuarios con la llegada de nuevas obras a la colección. Ciclos de larga demanda son de utilidad al momento de decidir adquisiciones porque reflejan que el género o el autor son buscados por los usuarios.

Socios de la biblioteca: Se obtuvo una vista gráfica que muestra cómo es la dinámica de las inscripciones como socios y cuál es el rango etario de aquellos que se asocian y cuáles son los departamentos de influencia de la biblioteca en San Juan. La Sociedad Franklin tiene socios en 15 de los 19 departamentos en los que se divide San Juan.

A los informes y gráficas generadas se suman otros relacionados a la saturación de la colección para identificar si hay áreas de la colección altamente densas e infrautilizadas; la estacionalidad por tema es otro informe que permite descubrir patrones de obras que tienen picos de préstamos en determinadas fechas, lo que colabora para planificar exhibiciones de las obras en las estanterías más visibles.

Seguramente se agregarán más análisis a los datos almacenados en DiGiBePe a partir de las solicitudes que se realicen por parte de los bibliotecarios durante la ejecución de proyecto en el 2025.

A fin de complementar estos datos extraídos del sistema de gestión bibliotecaria, se implementaron herramientas de recolección de información adicional:

- **Contador de gente**: Basado en un sensor infrarrojo, permite registrar la circulación de usuarios al ingresar y salir de la biblioteca, brindando datos detallados sobre los patrones de asistencia diarios.
- **Encuesta a usuarios en sala**: Aplicada con una muestra estadística basada en el *contador de gente*, permitió recoger información sobre la percepción de los servicios y la satisfacción general de los usuarios.

- **Software de control de uso de equipos:** Instalado para monitorear las horas de uso de los 8 equipos de acceso público disponibles en la biblioteca.

Estas acciones de recolección de datos han permitido a los bibliotecarios descubrir patrones de comportamiento de los usuarios en su paso por la biblioteca no solo a partir de la solicitud de obras de la colección. Por primera vez se puede graficar la gran cantidad de horas de uso de los equipos públicos instalados como herramientas de apoyo a la comunidad que no tiene acceso a estas herramientas (Pc, Tablet, Internet, etc.). Los resultados obtenidos de la encuesta realizada, permite a las autoridades de la biblioteca tomar decisiones basadas en datos para mejorar o reforzar sus servicios.

.

4. Formación de Recursos Humanos

Como parte del plan de ejecución del proyecto PDTS Datos Populares, se han impartido talleres de uso de *DiGiBePe*, *generación de informes SQL e introducción al uso de Tablas Dinámicas en Excel*. Todas estas capacitaciones fueron dirigidas a bibliotecarios de bibliotecas populares de San Juan.

De manera contínua se transfieren los conocimientos adquiridos a las cátedras de las carreras del Dpto. de Informática de la FCEFN y al equipo de trabajo que forma parte del Laboratorio de Bibliotecas Digitales y del Laboratorio de Sistemas Inteligentes para la Búsqueda de Conocimiento en Datos Masivos (Instituto de Informática, Resol. 02-2015).

5. Bibliografía

Estudio de FESABID "El valor económico y social de los servicios de información: bibliotecas" (2014). Recuperado de http://www.fesabid.org/wp-content/uploads/repositorio/fesabid-valor-economico-social-servicios-informacion-bibliotecas.pdf

Estudio sobre el retorno de la inversión de la Red de Bibliotecas Municipales de la provincia de Barcelona (2013). Recuperado de https://www.diba.cat/documents/16060163/183807107

Finch, Jannette (2016). Using Data Visualization to Examine an Academic Library Collection. College & Research Libraries, Noviembre, 165-178.

Olguin, Luis. (2016). DataViz en Bibliotecas. Recuperado de https://sedici.unlp.edu.ar/bitstream/handle/10915/61634/Documento_completo.pdf

Phetteplace. (2016). Data Visualization: A Guide to Visual Storytelling for Libraries. Edited by Lauren Magnuson, editor

OCLC Research. (2022, 9 de agosto). *Libraries support data-driven decision-making*. Hanging Together. https://hangingtogether.org/libraries-support-data-driven-decision-making/

· CGIV ·

Computación Gráfica, Imágenes y Visualización

ANÁLISIS VISUAL DE DATOS MULTIDIMENSIONALES

Ganuza, M. Luján[1,2,3] , Antonini, Antonella S.[1,2,3], Luque, Leandro E.[1,3], Selzer, Matías[1,2,3] , Herlein, Rodrigo N.[1,2], Larrea, Martín L.[1,2,3] , Tanzola, Juan E.[4], Asiain, Lucía[5], Ferracutti, Gabriela R.[4,5], Gargiulo, M. Florencia[4,5] , Bjerg, Ernesto A.[4,5], Castro, Silvia M.[1,2,3]

[1]Laboratorio de I+D en Visualización y Computación Gráfica (VyGLab) (UNS-CIC)

[2]Dpto. de Cs. e Ing. de la Computación, Universidad Nacional del Sur (DCIC-UNS)

{mlg, antonella.antonini, leandro.luque, matias.selzer}@cs.uns.edu.ar, rodrigonherlein@gmail.com, {mll, smc}@cs.uns.edu.ar

[3]ICIC, Instituto de Ciencias e Ingeniería de la Computación (UNS-CONICET)

[4]INGEOSUR, Instituto Geológico del Sur (UNS-CONICET)

[5]Dpto. de Geología, Universidad Nacional del Sur (DG-UNS)

{jetanzola,lasiain,ebjerg}@ingeosur-conicet.gob.ar, {gferrac,mfgargiulo}@uns.edu.ar

RESUMEN

La visualización de datos multidimensionales es fundamental para comprender las relaciones entre múltiples variables. Sin embargo, a mayor cantidad de dimensiones, más difícil resulta crear representaciones visuales efectivas que capturen todas las interrelaciones sin perder información ni generar oclusión visual. Si bien existen diversos métodos para visualizar datos multidimensionales, estos todavía presentan limitaciones que dificultan su escalabilidad y efectividad. En este contexto, nos enfocamos en mejorar las técnicas de visualización para datos multidimensionales con y sin pérdida de información. Para evaluar la efectividad de las técnicas propuestas, proponemos la implementación de estudios de seguimiento ocular que permiten analizar cómo los usuarios exploran e interpretan estas visualizaciones.

Palabras claves: Análisis Visual de Datos Multidimensionales, Visualización de Datos, Visualización sin Pérdida de Información.

CONTEXTO

Este trabajo se realiza en estrecha colaboración con investigadores del INGEOSUR-CONICET (Instituto Geológico del Sur), del Departamento de Geología de la Universidad Nacional del Sur, y el VyGLab (Laboratorio de Investigación y Desarrollo en Visualización y Computación Gráfica) del Departamento de Ciencias e Ingeniería de la Computación (DCIC-UNS), instituciones de investigación de reconocido prestigio tanto nacional como internacional.

1. INTRODUCCIÓN

La visualización de datos es una herramienta esencial para analizar y comprender los grandes volúmenes de información que se generan en diversas áreas científicas [Cle93, CMS99]. Sin embargo, cuando los datos tienen múltiples dimensiones, su visualización se vuelve particularmente desafiante. Mientras que los datos bidimensionales y tridimensionales pueden representarse de manera directa en nuestro mundo físico, la visualización de datos n-dimensionales requiere estrategias más sofisticadas que permitan considerar simultáneamente múltiples variables y sus interrelaciones. Se requieren herramientas de visualización avanzadas capaces de representar datos n-dimensionales en espacios 2D o 3D preservando la información original. Las técnicas tradicionales frecuentemente ofrecen solo vistas parciales de los datos debido a

limitaciones en el manejo de la dimensionalidad o el volumen, lo que obstaculiza la detección de relaciones relevantes entre las variables y dificulta el análisis integral del fenómeno estudiado.

Si bien existen diversas técnicas de visualización multidimensional que han demostrado ser efectivas para casos específicos, muchas de ellas solo logran mostrar vistas parciales de los datos, lo que dificulta detectar patrones y relaciones importantes entre las variables. Esto evidencia la necesidad de desarrollar nuevos métodos de visualización que puedan representar datos n-dimensionales en espacios 2D o 3D minimizando la pérdida de información y la oclusión visual. Tales métodos son particularmente valiosos ya que potencian las capacidades cognitivas humanas para descubrir patrones en datos multidimensionales [Kov18].

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El objetivo general de este trabajo se centra en contribuir al desarrollo de tecnologías y soluciones en torno al análisis visual de datos multidimensionales y su evaluación. En este contexto se plantean dos líneas principales de investigación:

El diseño y desarrollo de técnicas y herramientas para la visualización de datos multidimensionales

La visualización de datos multidimensionales enfrenta tres desafíos principales: la oclusión visual, la pérdida de información al proyectar datos n-dimensionales en espacios 2D o 3D, y la dificultad para crear representaciones efectivas y expresivas. Si bien la reducción de dimensionalidad es una técnica común en análisis visual, generalmente implica una pérdida irreversible de información, ya que los datos originales no pueden reconstruirse completamente a partir de su proyección reducida, excepto en casos muy específicos. Las técnicas tradicionales como scatterplots e histogramas solo ofrecen vistas parciales de la información, sin permitir una representación completa ni reversible de los datos multidimensionales.

Las soluciones actuales para estos desafíos son limitadas. Las coordenadas paralelas y radiales son los métodos más utilizados que preservan la información original, pero sufren problemas significativos de oclusión. Aunque las visualizaciones sin pérdida de información permiten comparar más atributos simultáneamente, el desarrollo de nuevos métodos en esta área es escaso.

En este contexto, nuestro trabajo busca avanzar en el desarrollo de técnicas de visualización tanto con pérdida controlada como sin pérdida de información, enfocándonos en datos multidimensionales temporales, espaciales y espacio-temporales. Las soluciones propuestas serán validadas utilizando datos geológicos, ya que el trabajo de laboratorio en geología representa un área prometedora para la aplicación de visualizaciones multidimensionales.

Evaluación de las nuevas técnicas diseñadas utilizando dispositivos de seguimiento ocular

La evaluación es un componente crucial en el desarrollo de técnicas de visualización, ya que permite validar su efectividad, identificar limitaciones y diseñar mejoras que potencien la cognición [FNS17]. Tradicionalmente, las evaluaciones se han centrado en métricas como tiempos de respuesta y tasas de error, complementadas con retroalimentación cualitativa de los usuarios [Duc17].

El seguimiento ocular representa un avance significativo en la evaluación de visualizaciones al registrar el comportamiento

espacio-temporal del movimiento ocular. Esta técnica registra el comportamiento espacio temporal del movimiento ocular proporcionando variables adicionales más allá de las medidas estándar como los tiempos de finalización y las tasas de error, registrando puntos de fijación y sus duraciones mientras el participante se concentra naturalmente en resolver la tarea. A diferencia de los métodos tradicionales que evalúan indirectamente a través de entrevistas y experimentos controlados, el seguimiento ocular proporciona datos directos sobre los patrones de atención visual. Esta riqueza de datos también presenta nuevos desafíos, requiriendo tecnología especializada para la captura de datos, algoritmos más sofisticados para su procesamiento y herramientas de visualización interactivas para su análisis.

3.RESULTADOS OBTENIDOS Y ESPERADOS

De las líneas de trabajo delineadas se han obtenido resultados parciales. Los miembros del equipo de investigación trabajan desde hace tiempo y de manera sostenida en el diseño y desarrollo de soluciones en torno a datos multidimensionales para distintos dominios de aplicación [SBG+18; LGA+21; SUS+21; AGC+22; ALG+23; KSB+22, GUL+23, BG24, DUG+24, LAG+24], y aplicado a ciencias geológicas en particular [GCF+12; GFG+14; GGF+15; GFG+17; AGF+21; AFG+23; ALG+23, ALF+24, ATA+24, FAA+24, LGB+24, TFA+24]. Adicionalmente, en colaboración con el Laboratorio de Desarrollo en Neurociencias Cognitivas (LDNC) del Departamento de Ingeniería Eléctrica y Computadoras (DIEC) perteneciente a la Universidad Nacional del Sur (UNS), se trabaja en la utilización de técnicas de seguimiento ocular y en el diseño y desarrollo de herramientas de visualización para datos provenientes de dispositivos de seguimiento ocular [LGC+21; LGC+22]. Finalmente, y con respecto a la evaluación de las técnicas propuestas, se participa activamente en el ámbito de la Verificación y Validación del Software aplicado a técnicas de visualización en particular [LSG+22; LUG23]. Las investigaciones que se vienen llevando a cabo en el grupo de trabajo tienen tienen aplicaciones relevantes en diversos campos de la geología, incluyendo la exploración de recursos naturales, la planificación de obras civiles y la gestión ambiental. La visualización de datos multidimensionales resulta particularmente valiosa para el trabajo de laboratorio geológico, ya que permite integrar múltiples fuentes de información: observaciones de campo, características microscópicas y macroscópicas de muestras rocosas, y resultados de análisis geoquímicos e isotópicos. Estas herramientas informáticas diseñadas e implementadas brindarán soporte a los geólogos para comprender mejor las interacciones entre procesos geológicos y la composición de diversos elementos (minerales, rocas, sedimentos, fluidos y gases) en contextos específicos, sea para caracterizar una región geológica o para abordar problemas de contaminación y remediación ambiental.

Debe señalarse que las tareas a desarrollar para alcanzar los objetivos planteados se llevarán a cabo en estrecha colaboración tanto con integrantes del INGEOSUR, como con actores del ámbito nacional e internacional.

4. FORMACION DE RECURSOS HUMANOS

Proyectos de Investigación

-PIBAA - CONICET (2872021010 0824CO) *Análisis Visual de Datos Multidimensionales sin Pérdida de Información*. Directora: Dra. M. Luján Ganuza.

-PGI 24/ZN42 *Análisis Visual de Datos Multidimensionales*. Directora: Dra. M. Luján Ganuza.

-***PGI 24/N059*** *Proyecto de investigación: Verificación y Validación de Software en la Industria 5.0.* Director: Dr. Martín L. Larrea.

Tesis en Desarrollo

-*Análisis Visual de Datos Multidimensionales*, Doctorado en Cs. de la Computación. Antonella S. Antonini. Directoras: Dra. Silvia Castro y Dra. M. Luján Ganuza.

-*Análisis visual de datos provenientes de registradores oculares*, Doctorado en Cs. de la Computación. Leandro Luque. Directoras: Dra. Silvia Castro y Dra. M. Luján Ganuza.

Becas

-Antonella S. Antonini. *Análisis Visual de Datos Multidimensionales.* Beca doctoral 2018 CONICET. Adjudicada a partir de abril de 2019. Directores: Dra. Silvia Castro - Dr. Ernesto Bjerg.

-Leandro Luque. *Análisis visual de datos provenientes de registradores oculares.* Beca doctoral 2018 CONICET. Adjudicada a partir de abril de 2019. Directores: Dra. Silvia Castro - Dr. Osvaldo Agamennoni.

5. BIBLIOGRAFÍA

[AFG+23] Antonini, A., Ferracutti, G., Gargiulo, F., Bjerg, E., Castro, S. & Ganuza, M. L. (2023) Análisis visual de datos multidimensionales de química mineral correspondientes a oxiespinelos de xenolitos del manto con Spinel Web. Caso de estudio. En Actas del 14° Congreso de Mineralogía, Petrología Ígnea y Metamórfica, y Metalogénesis - 14° MinMet y 5° PIMMA - Serie D, Publicación Especial 16 (pp. 36-42).

[AGC+22] Antonini, A. S., Ganuza, M. L. & Castro, S. M. (2022) VISUEL - A Web Dynamic Dashboard for Data Visualization. J. Comput. Sci. Technol. 22(1), 42-57.

[AGF+21] Antonini, A. S., Ganuza, M. L., Ferracutti, G., Gargiulo, M. F., Matković, K., Gröller, E., …Castro, S. M. (2021) Spinel web: an interactive web application for visualizing the chemical composition of spinel group minerals. Earth Sci. Informatics, 14(1), 521-528.

[ALF+24] Antonini, A. S., Luque, L., Ferracutti, G. R., Bjerg, E. A., Castro, S. M. & Ganuza, M. L. (2024). SpinelVA. A new perspective for the visual analysis and classification of spinel group minerals. Earth Science Informatics, 17, 3851-3861.

[ALG+23] Antonini, A. S., Luque, L., Ganuza, M. L. & Castro, S. M. (2023) Towards a Taxonomy for Non-Paired General Line Coordinates: A Comprehensive Survey. Int. J. Data Sci. Anal. 2023, (15), 133-158.

[ATA+24] Antonini, A., Tanzola, J., Asiain, L., Ferracutti, G., Castro, S., Bjerg, E. & Ganuza, M. L. (2024). Machine Learning model interpretability using SHAP values: application to Igneous Rock Classification task. Applied Computing and Geosciences, 23, 100178.

[BG24] Barrionuevo, M. & Ganuza, M. L. (2024). Improving Cryptocurrency Visual Analysis with CryptoVisualizer+. En Communications in Computer and Information Science - Computer Science – CACIC 2023 (Vol. 2123, pp. 107-118).

[Cle93] Cleveland, W. S. (1993) Visualizing Data. Hobart Press, 1993.

[CMS99] Card, S., Mackinlay, J., Shneiderman, B., (1999) Readings in Information Visualization – Using Vision to Think, Morgan Kaufmann.

[DUG+24] Diaz, G., Urribarri, D., Ganuza, M. L., Chesñevar, C. I., Estevez, E. & Maguitman, A. (2024). Polviz: Assessing Opinion Polarization in Social Media through Visual Analytics and Argumentation. ICEGOV'24: Proceedings of the 17th International Conference on Theory and Practice of Electronic Governance, 337-347.

[Duc17] Duchowski, T. (2017) Eye tracking: methodology theory and practice. Springer.

[FAA+24] Ferracutti, G. R., Asiain, L. M., Antonini, A., Tanzola, J. E. & Ganuza, M. L. (2024). OXYEMG: an application for determination of the oxyspinel group end-members based on electron microprobe analyses. European Journal of Mineralogy (EJM), (36), 87-98.

[FNS17] Fu, B., Noy, N. F., & Storey, M..A. (2017) Eye tracking the user experience–an evaluation of ontology visualization techniques. Semantic Web 8, 1, 23–41.

[GCF+12] Ganuza, M. L., Castro, S. M., Ferracutti, G., Bjerg, E. A., & Martig, S. (2012) Spinelviz: An

interactive 3d application for visualizing spinel group minerals. Comput. Geosci. 48, 50–56.

[GFG+14] Ganuza, M. L., Ferracutti, G., Gargiulo, M. F., Castro, S. M., Bjerg, E. A., Gröller, E., & Matkovic, K. (2014) The spinel explorer - interactive visual analysis of spinel group minerals. IEEE Trans. Vis. Comput. Graph. 20, 12, 1913–1922.

[GFG+17] Ganuza, M. L., Ferracutti, G., Gargiulo, M. F., Castro, S. M., Bjerg, E. A., Gröller, E., & Matkovic, K. (2017) Interactive visual categorization of spinel-group minerals. SCCG.

[GGF+15] Ganuza, M.L., Gargiulo, M.F., Ferracutti, G., Castro, S. M., Bjerg, E. A., Gröller, E., & Matkovic, K. (2015) Interactive semi-automatic categorization for spinel group minerals. IEEE VAST 2015, Chicago, IL, USA, October 25-30, pp. 197–198.

[GUL+23] Ganuza, M. L., Urribarri, D. K., Larrea, M. L. & Castro, S. M. (2023). UVAM: The Unified Visual Analytics Model. The Unified Visualization Model Revisited. En Cloud Computing, Big Data & Emerging Topics (pp. 189-203).

[Kov18] Kovalerchuk, B. (2018) Visual knowledge discovery and machine learning, vol. 144. Springer.

[KSB+22] Kuťák, D., Selzer, M. N., Byška, J., Ganuza, M. L., Barišić, I., Kozlíková, B. & Miao, H. (2022) Vivern–A Virtual Environment for Multiscale Visualization and Modeling of DNA Nanostructures. IEEE T Vis. Comput. Gr., 28(12), 4825-4838.

[LAG+24] Luque, L., Antonini, A., Ganuza, M. L. & Castro, S. M. (2024). GLC-Frame: A Framework and Library for Exploration of Multidimensional Data with General Line Coordinates. Journal of Computer Science and Technology, 24(1), e02.

[LGA+21] Luque, L. E., Ganuza, M. L., Antonini, A. S. & Castro, S. M. (2021) npGLC-Vis Library for Multidimensional Data Visualization. JCC-BD&ET.

[LGB+24] Luque, L., Ganuza, M. L., Bjerg, E. & Kovalerchuk, B. (2024). Visual Analysis of Spinels with General Line Coordinates. IV2024 - 28th International Conference Information Visualisation, 221-226.

[LGC+21] Luque, L., Ganuza, M. L., Castro, S. & Agamennoni, O. E. (2021) A visualization technique to support exploratory analysis of eye movements variables in reading. En XXXV Annual Meeting of the Argentinian Society for Neuroscience Research.

[LGC+22] Luque, L., Ganuza, M. L., Castro, S. M. & Agamennoni, O (2022). Visual analysis of eye movements during micro-stories reading. J. Vis., 25(5), 1085-1101.

[LSG+22] Larrea, M., Schiaffino, M., Ganuza, M. L. & Urribarri, D. K. (2022) A Testing Tool for Information Visualizations based on User Interactions. J. Comput. Sci. Technol. 22(1), 78-92.

[LUG23] Larrea, M. L., Urribarri, D. K. & Ganuza, M. L. (2023) New Testing Techniques to Evaluate the Quality of Information Visualization Implementations. LACCEI.

[SBG+18] Splechtna, R., Beham, M., Gracanin, D., Ganuza, M. L., Bühler, K., Pandžic, I. S., & Matkovic, K. (2018) Cross-table linking and brushing: interactive visual analysis of multiple tabular data sets. Vis. Comp.

[SUS+21] Sabando, M. V., Ulbrich, P., Selzer, M., Byška, J., Mičan, J., Ponzoni, I., M. L. Ganuza & Kozlíková, B. (2021) ChemVA: Interactive Visual Analysis of Chemical Compound Similarity in Virtual Screening. IEEE T Vis. Comput. Gr., 27(2), 891-901.

[TFA+24] Tanzola, J., Ferracutti, G., Asiain, L., Antonini, A. & Ganuza, M. L. (2024). Applied geochemistry for the discrimination between mafic and ultramafic rocks in Cu-Ni-PGE-bearing layered complexes: A case study at the la Jovita-Las Águilas belt, Sierra Grande de San Luis, Argentina. Journal of South American Earth Sciences, 134, 104755.

ANÁLISIS VISUAL DE DATOS MULTIDIMENSIONALES Y ESPACIO-TEMPORALES ASISTIDO POR MODELOS GENERATIVOS

Luque, Leandro E.[1,3], **Ganuza, M. Luján**[1,2,3], **Castro, Silvia M.**[1,2,3]
[1]Laboratorio de I+D en Visualización y Computación Gráfica (VyGLab) (UNS-CIC)
[2]Dpto. de Cs. e Ing. de la Computación, Universidad Nacional del Sur (DCIC-UNS)
[3]ICIC, Instituto de Ciencias e Ingeniería de la Computación (UNS-CONICET)
{ leandro.luque, mlg, smc}@cs.uns.edu.ar

RESUMEN

El análisis visual combina técnicas automatizadas con visualizaciones interactivas para entender y tomar decisiones sobre grandes conjuntos de datos complejos. La integración de modelos generativos de inteligencia artificial ha ampliado las capacidades de estas herramientas, permitiendo mejoras diversas tareas como el mapeo visual y la personalización automática. Este trabajo se centra en investigar la intersección entre análisis visual y modelos generativos para manipular datos multidimensionales y espacio-temporales en diversos dominios. El objetivo es desarrollar nuevas técnicas y herramientas que aprovechen el poder de razonamiento de los modelos generativos, adaptándose a las necesidades de expertos de diversas áreas.

Palabras claves: Análisis Visual de Datos, Datos Multidimensionales y Espacio-Temporales, Modelos Generativos

CONTEXTO

Esta línea de trabajo se realiza en el ámbito del VyGLab (Laboratorio de Investigación y Desarrollo en Visualización y Computación Gráfica) del Departamento de Ciencias e Ingeniería de la Computación (DCIC-UNS) en colaboración estrecha con el INGEOSUR (CONICET-UNS) y el Department of Computer Science (Central Washington University), instituciones de investigación de reconocido prestigio nacional e internacional.

1. INTRODUCCIÓN

En el último tiempo ha surgido una nueva área dentro del campo de la visualización denominada Inteligencia Artificial Generativa para Visualización (GenAI4Vis) [YHY+2024, BM+24] que busca integrar modelos generativos en diversas etapas del análisis visual [KAF+08]. De esta manera se puede distinguir un conjunto de tareas que emergen como consecuencia de dicha integración y son: la mejora de datos, la generación del mapeo visual, la personalización y las interacciones. La mejora de datos se encuentra vinculada a las primeras etapas del proceso de análisis visual y refiere al conjunto de tareas que buscan asegurar la completitud de los datos cuando sea necesario o mejorar la representación de los mismos. La generación del mapeo visual implica el diseño y desarrollo de algoritmos que permitan la generación automática de técnicas de visualización que cumplan con los criterios de diseño sin intervención de una persona. La personalización refiere a las mejoras estéticas que se le puede aplicar a la visualización teniendo en cuenta algún criterio que maximice la percepción por parte de los usuarios [VRS+22]. Por último, las interacciones es un aspecto clave donde se puede innovar, ya que el uso de modelos generativos da lugar a nuevas maneras de interactuar con las visualizaciones. Algunos ejemplos son: los mecanismos conversacionales donde es posible interactuar mediante el lenguaje natural con la

herramienta y obtener una retroalimentación constante; y la posibilidad de mejorar las interacciones tradicionales, como el zoom, de manera tal que permitan añadir más información vinculada al contexto.

Actualmente, existen algunas primeras versiones de herramientas que buscan hacer uso de los modelos de última generación para mejorar el proceso de análisis visual como LIDA [DIBIA+23], DASH [BS+24], DracoGPT [WGB+24], NL2VIS [LTL+21], Data2Vis [DD+19], DynaVis [VGI+24]. Sin embargo, sus principales desventajas radican en tres puntos: no soportan de forma integral todas las tareas definidas anteriormente, la potencia del modelo se encuentra vinculada a modelos privativos en la mayoría de los casos y en su intento de abarcar la mayor cantidad de dominios de aplicación posibles, las representaciones visuales que se obtienen como resultado son simples. La mayoría de las herramientas presentadas están enfocadas en la tarea de generación de mapeo visual, a excepción de DragoGPT [WGB+24] que busca extraer las decisiones de diseño subyacentes en múltiples técnicas de visualización.

Por otro lado, es importante destacar que la explosión del ecosistema de modelos generativos *open source* basados en lenguaje (LLM) como DeepSeek [GYZ+25], Qwen [YYZ+25], Llama [TMS+23], etc., permite la democratización del conocimiento y la posibilidad de poder construir soluciones de análisis visual basadas en IA a una fracción del costo técnico requerido y con un poder de expresividad y razonamiento cercano a los de los modelos más complejos empleados en la actualidad.

Finalmente, un aspecto importante referido al desarrollo de esta nueva área es la evaluación de las técnicas de análisis visual asistidas por los métodos generativos. Resulta de interés considerar diferentes métricas referidas a la aplicación de IA generativa en análisis visual como la exactitud y la fidelidad, alineamiento de intención (cuán efectivamente se traducen o materializan las intenciones del usuario), la robustez y la consistencia (la capacidad del modelo de ser aplicado en diversos dominios) y finalmente el bias y la ética (conceptos vinculados al uso que se le da al modelo generativo). Por otro lado, las técnicas de visualización generadas por los modelos de IA también son alcanzadas por métricas ya establecidos en el campo [BBK+18], que permiten medir la calidad de las representaciones y por lo tanto obtener un medida cuantitativa para refinar los resultados de forma iterativa.

Teniendo en cuenta los trabajos desarrollados por el equipo de trabajo [LGA+21, LGC+22, ALG+23, LAG+24, ALF+24, LGB+24] y la experiencia ganada en el campo del análisis visual, producto de la colaboración interdisciplinaria con otros grupos de investigación nacionales e internacionales de primer nivel, se propone extender y avanzar en el conocimiento sobre el análisis visual inteligente en el contexto de datos provenientes de dispositivos de seguimiento ocular (eye tracking), datos provenientes de las ciencias geológicas y datos multidimensionales en general.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El objetivo general es aportar al desarrollo y crecimiento del campo emergente de inteligencia artificial generativa aplicada al análisis visual y exploratorio de datos. Se hará hincapié sobre aquellos conjuntos de datos que poseen una naturaleza multidimensional y espacio-temporal. De esta manera se buscará desarrollar nuevos modelos, técnicas y/o herramientas que permitan aprovechar el poder de razonamiento que ofrecen los

modelos generativos y así construir herramientas de análisis visual inteligentes que se adapten a los requerimientos y tareas de los expertos. En este contexto se plantean las siguientes líneas de investigación:

-Redefinición, diseño y desarrollo del pipeline tradicional de análisis visual [GUL+23] que tenga por objeto principal a los modelos generativos. Esto implica diseñar técnicas de visualización que permitan incorporar en su definición la información obtenida de estos modelos. Además, se explorarán nuevas formas de especificar las visualizaciones generadas para favorecer su comparación, intercambio y modificación. Se deberán diseñar e incorporar interacciones nuevas que faciliten la retroalimentación entre el modelo, la técnica de visualización y el usuario.

-Mejora de técnicas de análisis visual multidimensional, basadas en el desarrollo de trabajos previos [LGA+21, ALG+23, LAG+24] donde será necesario explorar cómo incorporar y extender conceptos como la dimensión temporal y nuevos algoritmos de renderizado

-Actualización de las distintas métricas de evaluación propuestas en la literatura para la evaluación de las nuevas técnicas de análisis visual de datos guiadas por modelos generativos, a fin de contar con métodos y criterios que acompañen el ritmo de crecimiento de esta nueva área.

Estas líneas de investigación fueron establecidas en función de un conjunto de hipótesis específicas que se espera alcanzar durante el transcurso de la investigación:

- Explorar e investigar las diferentes maneras que existan de automatizar y/o mejorar ciertas etapas del proceso de análisis visual. Gracias al avance de los modelos generativos, se considera que el estudio vinculado a la generación, sugerencia y mejora de visualizaciones constituye un avance significativo en la manera en que los usuarios expertos se vinculan con las herramientas de análisis visual para la obtención de resultados.
-La publicación constante de modelos generativos *open source* que pueden ser verificados y adaptados, permite construir herramientas que se adapten a los requerimientos tecnológicos con los que se dispone.

3.RESULTADOS OBTENIDOS Y ESPERADOS

Estas líneas de investigación han sido desarrolladas en el último tiempo por el grupo de investigación del VyGLab en conjunto con profesionales de las Ciencias Geológicas y Neurociencias, trabajando fuertemente en temas vinculados con la visualización de datos aplicada a diversos dominios de aplicación. Los trabajos [LGC+20, LGC+22] reflejan el aporte en el campo del análisis visual de datos provenientes de eye tracking, al proponer técnicas para la representación visual de series de datos espacio-temporales correspondientes a la lectura, en colaboración con expertos en neurociencias pertenecientes al Laboratorio de Desarrollo en Neurociencias Cognitivas (LDNC). Durante el desarrollo de estos trabajos se identificó la necesidad de explorar e investigar un nuevo conjunto de técnicas para representar datos multidimensionales, dando lugar a una serie de trabajos basados en Coordenadas Generales de Líneas (GLC). Estos se encuentran organizados en tres pilares fundamentales que son: la definición de una taxonomía y relevamiento de la literatura correspondiente a un subconjunto de interés de las GLC [ALG+23], el diseño y desarrollo de herramientas que permitan emplear dichas técnicas y evaluar cómo las interacciones asociadas pueden ser empleadas para descubrir patrones de interés [LGA+21,

LAG+24] y finalmente la aplicación práctica de las GLC sobre un problema real correspondiente al dominio de la geología [LGB+24]. Este último trabajo ha sido realizado en conjunto con el Dr. Boris Kovalerchuk (CWU), que ha sido el creador y propulsor de las GLC como alternativa para representar datos sin pérdida de información, y ha derivado en un vínculo de colaboración académica internacional sobre desarrollos basados en GLC que persiste hasta la actualidad. Por otro lado, el trabajo [ALF+24] realizado junto con profesionales del INGEOSUR (CONICET-UNS) ha sido el primer acercamiento a al desarrollo de soluciones de análisis visual asistido por inteligencia artificial, al combinar técnicas de análisis visual con métodos de machine learning de clasificación y reducción de dimensionalidad para la caracterización de ambientes tectónicos.

4. FORMACION DE RECURSOS HUMANOS

Proyectos de Investigación

-PIBAA - CONICET (2872021010 0824CO) *Análisis Visual de Datos Multidimensionales sin Pérdida de Información*. Directora: Dra. M. Luján Ganuza.

-PGI 24/ZN42 *Análisis Visual de Datos Multidimensionales*. Directora: Dra. M. Luján Ganuza.

-PGI 24/ZN44 *Visualización Comparativa: un enfoque desde la visualización de información para la comparación de datos*. Directora: Dra. Dana Urribarri.

Tesis en Desarrollo

-*Análisis visual de datos provenientes de registradores oculares*, Doctorado en Cs. de la Computación. Leandro Luque. Directoras: Dra. Silvia Castro y la Dra. M. Luján Ganuza.

Becas

-Leandro Luque. *Análisis visual de datos provenientes de registradores oculares*. Beca doctoral 2018 CONICET. Adjudicada a partir de abril de 2019. Directores: Dra. Silvia Castro - Dr. Osvaldo Agamennoni.

5. BIBLIOGRAFÍA

[ALG+23] Antonini, A. S., Luque, L., Ganuza, M. L., & Castro, S. M. (2023). Toward a taxonomy for 2D non-paired General Line Coordinates: a comprehensive survey. International Journal of Data Science and Analytics, 15(2), 133-158.

[ALF+24] Antonini, A. S., Luque, L., Ferracutti, G. R., Bjerg, E. A., Castro, S. M., & Ganuza, M. L. (2024). SpinelVA. A new perspective for the visual analysis and classification of spinel group minerals. Earth Science Informatics, 17(4), 3851-3861.

[BS+24] Bromley, D., & Setlur, V. (2024). DASH: A Bimodal Data Exploration Tool for Interactive Text and Visualizations. In 2024 IEEE Visualization and Visual Analytics (VIS) (pp. 256-260). IEEE.

[BM+24] Basole, R. C., & Major, T. (2024). Generative AI for Visualization: Opportunities and Challenges. IEEE Computer Graphics and Applications, 44(2), 55-64.

[DIBIA+23] Dibia, V. (2023). LIDA: A Tool for Automatic Generation of Grammar-Agnostic Visualizations and Infographics using Large Language Models. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations), pages 113–126, Toronto, Canada. Association for Computational Linguistics.

[DD+19] Dibia, V., & Demiralp, Ç. (2019). Data2vis: Automatic generation of data visualizations using sequence-to-sequence recurrent neural networks. IEEE computer graphics and applications, 39(5), 33-46.

[GUL+23] Ganuza, M. L., Urribarri, D. K., Larrea, M. L., & Castro, S. M. (2023, June). UVAM: The Unified Visual Analytics Model. The Unified Visualization Model Revisited. In Conference on Cloud Computing, Big Data & Emerging Topics (pp. 189-203). Cham: Springer Nature Switzerland.

[GYZ+25] Guo, D., Yang, D., Zhang, H., Song, J., Zhang, R., Xu, R., ... & He, Y. (2025). Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. arXiv preprint arXiv:2501.12948.

[KAF+08] Keim, D., Andrienko, G., Fekete, J. D., Görg, C., Kohlhammer, J., & Melançon, G. (2008). Visual analytics: Definition, process, and challenges (pp. 154-175). Springer Berlin Heidelberg.

[LGC+20] Luque, L., Ganuza, M. L., Castro, S. M., & Agamennoni, O. E. (2020). A visualization technique to support exploratory analysis of eye movements' variables in reading. In SAN 2020 e-book del XXXV Annual Meeting of the Argentinian Society for Neuroscience Research p. 353, Congreso Virtual

[LGA+21] Luque, L. E., Ganuza, M. L., Antonini, A. S., & Castro, S. M. (2021). npGLC-Vis library for multidimensional data visualization. In Conference on Cloud Computing, Big Data & Emerging Topics (pp. 188-202). Cham: Springer International Publishing.

[LGC+22] Luque, L., Ganuza, M. L., Castro, S. M., & Agamennoni, O. E. (2022). Visual analysis of eye movements during micro-stories reading. Journal of Visualization, 25(5), 1085-1101.

[LAG+24] Luque, L. E., Antonini, A. S., Ganuza, M. L., & Castro, S. M. (2024). GLC-Frame: A Framework and Library for Exploration of Multidimensional Data with General Line Coordinates. Journal of Computer Science and Technology, 24(1), 14-18.

[LGB+24] Luque, L., Ganuza, M. L., Bjerg, E., & Kovalerchuk, B. (2024). Visual Analysis of Spinels with General Line Coordinates. In 2024 28th International Conference Information Visualisation (IV) (pp. 221-226). IEEE.

[LTL+21] Luo, Y., Tang, N., Li, G., Chai, C., Li, W., & Qin, X. (2021). Synthesizing natural language to visualization (NL2VIS) benchmarks from NL2SQL benchmarks. In Proceedings of the 2021 International Conference on Management of Data (pp. 1235-1247).

[BBK+18] Behrisch, M., Blumenschein, M., Kim, N. W., Shao, L., El‐Assady, M., Fuchs, J., ... & Keim, D. A. (2018). Quality metrics for information visualization. In Computer Graphics Forum (Vol. 37, No. 3, pp. 625-662).

[TMS+23] Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., ... & Scialom, T. (2023). Llama 2: Open foundation and fine-tuned chat models. arXiv preprint arXiv:2307.09288.

[VGI+24] Vaithilingam, P., Glassman, E. L., Inala, J. P., & Wang, C. (2024). DynaVis: Dynamically Synthesized UI Widgets for Visualization Editing. In Proceedings of the CHI Conference on Human Factors in Computing Systems (pp. 1-17).

[VRS+22] Vij, R., Raj, R., Singhal, M., Tanwar, M., & Bedathur, S. (2022). Vizai: Selecting accurate visualizations of numerical data. In Proceedings of the 5th Joint International Conference on Data Science & Management of Data (9th ACM IKDD CODS and 27th COMAD) (pp. 28-36).

[WGB+24] Wang, H. W., Gordon, M., Battle, L., & Heer, J. (2024). DracoGPT: Extracting Visualization Design Preferences from Large Language Models. IEEE Transactions on Visualization and Computer Graphics.

[YYZ+25] Yang, A., Yang, B., Zhang, B., Hui, B., Zheng, B., Yu, B., ... & Qiu, Z. (2024). Qwen2. 5 technical report. arXiv preprint arXiv:2412.15115.

[YHY+24] Ye, Y., Hao, J., Hou, Y., Wang, Z., Xiao, S., Luo, Y., & Zeng, W. (2024). Generative ai for visualization: State of the art and future directions. Visual Informatics.

Análisis Visual y Tecnologías Inmersivas para la Enseñanza en Ciencias de los Materiales

Facundo Bernardini[1], Germán Prieto[2], María Luján Ganuza[3], Matías N. Selzer[3]

[(1)]Dpto. de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur (DCIC-UNS)
facubernardini@hotmail.com

[(2)]Centro de Desarrollo en Tribología y Tecnología de Superficies, Departamento de Ingeniería UNS, IFISUR, (UNS-CONICET).
german.prieto@uns.edu.ar

[(3)] Laboratorio de I+D en Visualización y Computación Gráfica
Dpto. de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur (DCIC-UNS)
Instituto de Ciencias e Ingeniería de la Computación (UNS-CONICET)

RESUMEN

Esta línea de investigación se centra en el diseño y desarrollo de soluciones de visualización aplicadas a ciencia de los materiales con fines educativos, aprovechando las tecnologías inmersivas de bajo costo accesible a través de dispositivos móviles. El objetivo principal es crear una plataforma integral que combine técnicas de visualización y tecnologías inmersivas con el fin de proporcionar experiencias de aprendizaje inmersivas y efectivas en el campo de las ciencias de los materiales, en particular en el estudio de estructuras cristalinas y diagramas de fases binarios. Los resultados obtenidos serán difundidos a través de publicaciones académicas y se proyecta la utilización del framework en materias introductorias a la ciencia de los materiales de la Universidad Nacional del Sur.

Palabras Clave: Aprendizaje Inmersivo, Visualización de Datos, Tecnologías Inmersivas, Laboratorio Virtual, Cristalografía

CONTEXTO

Esta línea de investigación se lleva a cabo en el Departamento de Ciencias e Ingeniería de la Computación (DCIC-UNS). Trabajamos en estrecha colaboración con docentes responsables de materias introductorias a la ciencia de los materiales de la Universidad Nacional del Sur. Además, contamos con investigadores del Laboratorio de Investigación y Desarrollo en Visualización y Computación Gráfica (VyGLab) de la Universidad Nacional del Sur.

1. INTRODUCCIÓN

La visualización de datos es cada vez más importante en la educación, ya que permite una mejor comprensión de los temas a través de representaciones gráficas [FL18]. Los métodos de visualización interactiva son clave para simplificar y transmitir información significativa sobre sistemas complejos, resultando extremadamente útiles para todo tipo de usuarios.

En las ciencias de los materiales, comprender las estructuras cristalinas y la disposición atómica/molecular en los sólidos cristalinos es crucial para entender las propiedades de los materiales [CCM+23]. Sin embargo, la enseñanza y aprendizaje de estas estructuras presenta dificultades de visualización espacial, al tener que recrear mentalmente distribuciones espaciales complejas de átomos y celdas unitarias [FME18]. Muchos estudiantes encuentran difícil comprender estos conceptos cuando solo se utilizan representaciones gráficas 2D, lo cual dificulta la enseñanza basada exclusivamente en esos métodos.

En las últimas décadas, diversas técnicas de visualización han demostrado su utilidad para la enseñanza de estructuras cristalinas. Los enfoques actuales incluyen bocetos 2D [S76], animaciones [S08], impresiones 3D [RVC15, SV14], visualizaciones interactivas donde los estudiantes manipulan modelos 3D de cristales [WPT12, GJB+20], y diversos sitios web y/o programas que ofrecen visualizaciones interactivas de estructuras cristalinas y celdas unitarias. Estas técnicas buscan facilitar la comprensión de las complejas disposiciones atómicas/moleculares en los sólidos cristalinos mediante representaciones visuales e

interactivas más allá de los gráficos 2D tradicionales.

Por otra parte, los avances tecnológicos han impulsado el desarrollo de tecnologías inmersivas como la realidad virtual (RV) y realidad aumentada (RA), dando lugar a los laboratorios virtuales. Las tecnologías inmersivas combinan la realidad con objetos virtuales (RA) o sumergen completamente al usuario en mundos totalmente virtuales (RV). Aunque se suele pensar que estas tecnologías se utilizan principalmente para videojuegos, existen muchas aplicaciones en diversos campos, como la educación [CP23, HLL16, I+15, C16], el entrenamiento militar y médico [LPB+17, LLR+16, PG14], la medicina [HKL14], el turismo, entre otras [GDN+17].

Dispositivos de RV y RA de alta gama proveen una experiencia sumamente inmersiva, pero su alto costo los hace inaccesibles para muchos usuarios. Por este motivo las tecnologías inmersivas de bajo costo (aquellas que utilizan el propio teléfono celular del usuario) se han vuelto populares ya que proveen una alternativa económica y práctica [ASR+20]. Existen numerosos estudios con el objetivo de proporcionar entornos inmersivos de bajo costo [HK16, AMK+18, RPM+18, P+19]. En otras palabras, el alto costo de dispositivos es uno de los desafíos críticos en este tipo de tecnologías y debe tenerse en cuenta en futuros estudios [BSV01, CB19]. En los últimos años, las tecnologías inmersivas han revolucionado la educación al ofrecer experiencias inmersivas y dinámicas que facilitan la comprensión de conceptos abstractos [TBP23, JM23]. Estas tecnologías permiten la simulación de entornos complejos, la visualización interactiva de información y el aprendizaje personalizado, lo que mejora la retención de conocimiento y amplía el acceso a la educación [OHM23].

Varias propuestas académicas han explorado el uso de tecnologías inmersivas para enseñar cristalografía [EVD+20], demostrando que estas herramientas enriquecen la educación al proporcionar experiencias más inmersivas y prácticas. Por ejemplo, los estudiantes pueden explorar estructuras moleculares en entornos virtuales tridimensionales, lo que les ayuda a comprender cómo la estructura afecta las propiedades del material. También pueden realizar experimentos virtuales para observar cómo diferentes procesos afectan las propiedades de los materiales, sin los costos y riesgos de los experimentos reales. Esto fomenta la experimentación y la exploración, ayudando a los estudiantes a comprender mejor la ciencia de materiales y a desarrollar habilidades prácticas de manera segura y efectiva.

En este contexto, se plantea el diseño y desarrollo de soluciones que utilicen visualización para asistir en el área de ciencia de materiales con fines educativos, aprovechando las tecnologías inmersivas de bajo costo accesible a través de teléfonos celulares.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Visualización y manipulación de procesos de cristalización y formación de aleaciones.

Esta línea de investigación busca diseñar una plataforma visual y dinámica que permita estudiar los procesos complejos de cristalización desde una perspectiva innovadora y personalizable. El objetivo es crear simulaciones interactivas que permitan al usuario modificar en tiempo real las condiciones iniciales, como la cantidad de elementos en la aleación y las cantidades relativas de cada uno de ellos. Con esto, se podrá observar cómo estos factores afectan la estructura cristalina, principalmente la formación de defectos (vacancias, dislocaciones), bordes de grano, etc.

Este módulo permitirá explorar la formación de aleaciones, con un enfoque particular en la microestructura resultante, la interacción entre diferentes fases y la capacidad de predecir cómo estas aleaciones podrían comportarse bajo diferentes solicitaciones.

Interacción dinámica con celdas unitarias y sistemas de deslizamiento.

Esta línea de investigación tiene como objetivo crear un entorno de simulación avanzado en el que los usuarios puedan interactuar con celdas unitarias y visualizar las estructuras atómicas en 3D. La idea es proporcionar una herramienta de modelado que permita ver de manera detallada cómo se disponen

los átomos en las celdas unitarias. Uno de los enfoques clave será la visualización dinámica de los sistemas de deslizamiento, que son fundamentales para entender cómo los materiales responden a la deformación. Mediante una interfaz interactiva, los estudiantes podrán observar cómo las dislocaciones y los movimientos atómicos ocurren en tiempo real dentro de la estructura cristalina. Además, se dará especial atención a los espacios intersticiales, que son cruciales para el estudio de las propiedades de materiales como la difusión de átomos y la formación de defectos. Esta línea permitirá avanzar en la comprensión de las propiedades mecánicas y termodinámicas de los materiales a nivel atómico.

Evolución de fases en aleaciones y visualización de diagramas binarios.

El objetivo de esta línea de investigación es desarrollar una herramienta que permita la visualización de la evolución de fases en aleaciones metálicas, con un enfoque especial en los diagramas binarios de fases. Esta plataforma interactiva permitirá a los usuarios seleccionar distintos componentes de aleaciones y simular su enfriamiento, observando cómo las fases se desarrollan y cambian a lo largo del proceso. Además, se incluirá una visualización dinámica que permita no solo ver las fases a medida que se forman, sino también interactuar con ellas, alterando variables y observando su impacto en la estructura final.

3. RESULTADOS OBTENIDOS Y ESPERADOS

Hasta el momento, se ha avanzado en el desarrollo de una plataforma interactiva que permite la visualización y manipulación de estructuras cristalinas y diagramas de fases binarios mediante tecnologías inmersivas de bajo costo. Se han implementado modelos tridimensionales interactivos accesibles desde dispositivos móviles, lo que facilita la exploración y comprensión de configuraciones atómicas complejas en un entorno educativo intuitivo.

Entre los resultados obtenidos, se destacan:

- **Implementación de visualización interactiva:** Desarrollo de modelos 3D de celdas unitarias y diagramas de fases que permiten la manipulación por parte de los estudiantes.
- **Uso de realidad aumentada (RA):** Integración de una aplicación móvil que superpone modelos cristalográficos en el entorno físico, mejorando la comprensión espacial de las estructuras atómicas.

En cuanto a los resultados esperados, se proyecta:

- **Optimización del framework:** Ampliación de la biblioteca de modelos cristalográficos y diagramas de fases, con simulaciones dinámicas que permitan visualizar la evolución de estructuras en función de variables como temperatura y composición.
- **Expansión de la plataforma educativa:** Integración de nuevas funcionalidades que permitan la personalización de los modelos y su uso en distintas asignaturas de ciencia de materiales.
- **Evaluación pedagógica sistemática:** Realización de estudios comparativos para medir la efectividad del sistema en términos de aprendizaje y retención de conceptos.
- **Publicaciones científicas y difusión:** Presentación de los avances en congresos y revistas académicas, así como la implementación de la plataforma en cursos de grado en la Universidad Nacional del Sur.

Estos desarrollos contribuirán a la enseñanza interactiva en ciencias de los materiales, promoviendo metodologías innovadoras que potencien la comprensión y el aprendizaje de conceptos complejos a través de tecnologías inmersivas accesibles.

Este proyecto se está realizando en colaboración con docentes responsables de materias introductorias a la ciencia de los materiales de la Universidad Nacional del Sur, siendo el objetivo final producir una herramienta innovadora, para ser utilizada y evaluada por los alumnos en el aula.

4. FORMACIÓN DE RECURSOS HUMANOS

A continuación, se detallan los proyectos de investigación vinculados con las líneas de investigación presentadas y las becas adquiridas en centros de investigación científica.

Proyectos vigentes

- PGI SeCyT-UNS (24/ZN43) "*Tecnologías Inmersivas y Visualización Situada aplicadas a Geociencias"*. Director: Matias N. Selzer.
- PIBAA - CONICET (2872021010 0824CO) *"Análisis Visual de Datos Multidimensionales sin Pérdida de Información"*. Directora: Dra. M. Luján Ganuza.
- PGI SeCyT-UNS (24/ZN42) *"Análisis Visual de Datos Multidimensionales"*. Directora: María Luján Ganuza.
- PGI SeCyT-UNS (24/ZJ53) *"Generación y evaluación tribológica de recubrimientos blandos compuestos basados en nanotubos de carbono y sulfuros metálicos"*. Director: Germán Prieto.

Proyecto Final de Carrera en Desarrollo

• Facundo Bernardini *"Análisis Visual y Tecnologías Inmersivas Aplicadas a la Enseñanza Interactiva en Ciencias de los Materiales", Ingeniería en Computación*. Director: Dr. Matías Selzer, Codirectora: Dra. María Luján Ganuza. Colaborador: Dr. Germán Prieto.

5. BIBLIOGRAFÍA

[AMK+18] Ambron E., Miller A., Kuchenbecker K. J., Buxbaum L. J., and Coslett H., "Immersive low-cost virtual reality treatment for phantom limb pain: Evidence from two cases," Frontiers in neurology, vol. 9, p. 67, 2018.

[ASR+20] Asadzadeh, A., Samad-Soltani, T., Rezaei-Hachesu, P., & Salahzadeh, Z. (2020, April). Low-cost interactive device for virtual reality. In 2020 6th international conference on Web research (ICWR) (pp. 38-42). IEEE.

[BSV01] Belleman R. G., Stolk B., and de Vries R., "Immersive virtual reality on commodity hardware," in Proceedings of the 7th annual conference of the Advanced School for Computing and Imaging, 2001, vol. 7, pp. 297-304.

[CCM+23] Caro, V., Carter, B. A., Millunchick, J., & Reeves, S. (2023). Teaching crystal structures in undergraduate courses: a systematic review from a disciplinary literacy perspective. Chemistry Education Research and Practice, 24(2), 394-406.

[CB19] Checa D. and Bustillo A., "A review of immersive virtual reality serious games to enhance learning and training," Multimedia Tools and Applications, pp. 1-27, 2019.

[CP23] Cho, Y., & Park, K. S. (2023). Designing immersive virtual reality simulation for environmental science education. Electronics, 12(2), 315.

[C16] Choi, D.W., 2016, January. Emerging Tools and Applications of Virtual Reality in Education, IGI Global.

[EVD+20] Extremera, J., Vergara, D., Dávila, L. P., & Rubio, M. P. (2020). Virtual and augmented reality environments to learn the fundamentals of crystallography. Crystals, 10(6), 456.

[FL18] Fırat, E. E., & Laramee, R. S. (2018). Towards a survey of interactive visualization for education. Computer Graphics and Visual Computing, CGVC 2018, 91-101.

[FME18] Fogarty, J., McCormick, J., & El-Tawil, S. (2018). Improving student understanding of complex spatial arrangements with virtual reality. Journal of Professional Issues in Engineering Education and Practice, 144(2), 04017013.

[GDN+17] Grandi, J.G., Debarba, H.G., Nedel, L., and Maciel, A. (2017, May). Design and evaluation of a handheld-based 3d user interface for collaborative object manipulation. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (pp. 5881-5891).

[GJB+20] Gruber, C., James, A., Berchtold, J. T., Wood, Z. J., Scott, G. E., & Alghoul, Z. (2020). Interactive unit cell visualization tool for crystal lattice structures.

[HKL14] Henderson, A., Korner-Bitensky, N., Levin, M., 2014. Virtual reality in stroke rehabilitation: a systematic review of its effectiveness for upper limb motor recovery. Topics in stroke rehabilitation.

[HK16] Hilfert T. and König M., "Low-cost virtual reality environment for engineering and construction," Visualization in Engineering, vol. 4, no. 1, p. 2, 2016.

[HLL16] Huang, H.-M., Liaw, S.-S., Lai, C.-M., 2016. Exploring learner acceptance of the use of virtual reality in medical education: a case study of desktop and projection-based display systems. Interactive Learning Environments, 24(1), 3-19.

[I+15] Ibayashi, H., et al., 2015. Dollhouse VR: a multi-view, multi-user collaborative design workspace with VR technology. SIGGRAPH Asia 2015 Emerging Technologies. 1-2, 2015.

[JM23] Jiawei, W., & Mokmin, N. A. M. (2023). Virtual reality technology in art education with visual communication design in higher education: a systematic literature review. Education and Information Technologies, 28(11), 15125-15143.

[LPB+17] Lacoche, J., Pallamin, N., Boggini, T., and Royan, J. (2017, November). Collaborators awareness for user cohabitation in co-located collaborative virtual environments. In Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology (pp. 1-9).

[LLR+16] Le Chénéchal, M., Lacoche, J., Royan, J., Duval, T., Gouranton, V., and Arnaldi, B. (2016, March). When the giant meets the ant an asymmetric approach for collaborative and concurrent object manipulation in a multi-scale environment. In 2016 IEEE Third VR International Workshop on Collaborative Virtual Environments (3DCVE) (pp. 18-22). IEEE.

[OHM23] Oje, A. V., Hunsu, N. J., & May, D. (2023). Virtual reality assisted engineering education: A multimedia learning perspective. Computers & Education: X Reality, 3, 100033.

[P+19] Patel P. et al., "Low-Cost Virtual Reality Headsets Reduce Perceived Pain in Healthy Adults: A Multicenter Randomized Crossover Trial," Games for Health Journal, 2019.

[PG14] Palter, V. N., Grantcharov, T. P., 2014. Individualized deliberate practice on a virtual reality simulator improves technical performance of surgical novices in the operating room: a randomized controlled trial. Annals of surgery, 259(3), 443-448.

[RVC15] Rodenbough, P. P., Vanti, W. B., & Chan, S. W. (2015). 3D-printing crystallographic unit cells for learning materials science and engineering. Journal of Chemical Education, 92(11), 1960-1962.

[RPM+18] Román-Ibáñez V., Pujol-López F. A., Mora-Mora H., Pertegal-Felices M. L., and Jimeno-Morenilla A., "A low-cost immersive virtual reality system for teaching robotic manipulators programming," Sustainability, vol. 10, no. 4, p. 1102, 2018.

[SV14] Scalfani, V. F., & Vaid, T. P. (2014). 3D printed molecules and extended solid models for teaching symmetry and point groups. Journal of Chemical Education, 91(8), 1174-1180.

[S08] Soper, L. M., Hawkins, J. A., Rittenhouse, J. L., & Rittenhouse, R. C. (2008). Use of the Primitive Unit Cell in Understanding Subtle Features of the Cubic Close-Packed Structure. Journal of Chemical Education, 85(1), 90.

[S76] Suchow, L. (1976). Other views of unit cells. Journal of Chemical Education, 53(4), 226. doi:10.1021/ed053p226

[TBP23] Tiwari, C. K., Bhaskar, P., & Pal, A. (2023). Prospects of augmented reality and virtual reality for online education: a scientometric view. International Journal of Educational Management, 37(5), 1042-1066.

[WPT12] Wilson, C. C., Parkin, A., & Thomas, L. H. (2012). Frontiers of crystallography: A project-based research-led learning exercise. Journal of Chemical Education, 89(1), 34-37.

Aplicación de TIC para el Monitoreo y Gestión del Ganado en áreas extensas

Lazarte, Ivanna[1,2]; Aranda, Marcos[1,2]; Beltramini, Paola[2]; Moreno, Juan Pablo[1,2]; Villagrán, Daniel[2]; Cano, Eduardo[2]; Gallina, Sergio[2]

[1] LaTICs/Facultad de Tecnología y Ciencias Aplicadas/Universidad Nacional de Catamarca
[2] Grupo IoT/Facultad de Tecnología y Ciencias Aplicadas/Universidad Nacional de Catamarca
Maximio Victoria Nº 55 - C.P: 4700 - San Fernando del Valle de Catamarca.
Tel: +54 (03834) 435112

ilazarte@tecno.unca.edu.ar, maranda@tecno.unca.edu.ar, pbeltramini@tecno.unca.edu.ar

RESUMEN

El presente trabajo se enmarca en el paradigma Agro 4.0 y describe una propuesta de investigación para el desarrollo de un sistema integral basado en tecnologías de la información y las comunicaciones (TIC) para el monitoreo y análisis del comportamiento del ganado en sistemas extensivos del Noroeste Argentino (NOA). A través del uso de dispositivos IoT, se recopilarán datos georreferenciados sobre la ubicación y desplazamiento del ganado, integrándolos con imágenes satelitales y de drones para evaluar la disponibilidad y calidad de los pastizales y también el conteo automático del ganado.

El proyecto se estructura en seis fases: monitoreo georreferenciado del ganado mediante collares GPS, selección de índices de vegetación para evaluar los pastizales, desarrollo de herramientas de visualización de zonas de pastoreo, estudio de técnicas de procesamiento de imágenes para analizar el comportamiento animal, aplicación de redes neuronales para el conteo automático de ganado y difusión de los resultados obtenidos.

Se espera que la integración de estas tecnologías facilite la toma de decisiones basada en datos en tiempo real, optimizando la gestión ganadera y promoviendo la digitalización del sector. Este enfoque contribuirá a mejorar la eficiencia productiva y el uso sostenible de los recursos en la ganadería extensiva del NOA.

***Palabras clave*:** IoT, ganadería extensiva, imágenes satelitales, machine learning

CONTEXTO

Este trabajo de investigación se lleva a cabo en conjunto con el Grupo IoT y el Laboratorio de Tecnologías de Información y las Comunicaciones (LaTIC's), de la Facultad de Tecnología de la UNCA y se encuentra enmarcado dentro del Proyecto de I+D "Innovación Digital en la Ganadería Extensiva del NOA: Aplicación de TIC para el Monitoreo y Gestión del Ganado". El proyecto fue aprobado en la Convocatoria a Proyectos de Investigación Científico Tecnológica Interdisciplinaria de Impacto Provincial y Regional (PICTI - IPR). Se enmarca en la línea de investigación "Aplicación de Tecnologías de la Información y Comunicación (TIC) e Inteligencia Artificial en la Agroindustria", con enfoque en la digitalización y automatización de procesos en la ganadería extensiva mediante IoT, procesamiento de imágenes satelitales y análisis de datos.

1. INTRODUCCIÓN

El desafío actual y futuro de la ganadería es alcanzar una mayor eficiencia productiva, previsible y sostenible, asociada a la aplicación de tecnologías que permitan monitorear el ganado y la infraestructura en tiempo real, facilitando la anticipación de riesgos, el ahorro de tiempo y la optimización de costos. En los

sistemas ganaderos intensivos, el monitoreo continuo del ganado representa una dificultad clave, ya que las prácticas tradicionales, como el conteo e inspección manual de animales, requieren mucho tiempo y mano de obra. Estas prácticas convencionales incluyen la medición del peso corporal mediante básculas industriales, las cuales, además de requerir una inversión considerable, deben instalarse en lugares específicos y son costosas de mantener. Asimismo, el uso de dispositivos manuales como cintas métricas y otros instrumentos para medir características morfológicas del ganado requiere la sujeción física de los animales, lo cual no solo es estresante para ellos, sino que también implica riesgos para los trabajadores. La evaluación de la condición corporal, por su parte, depende de la experiencia del observador, ya que es una metodología subjetiva que evalúa visual y manualmente la grasa y el estado nutricional del animal, lo que puede llevar a inconsistencias en los resultados. Por otro lado, la detección de problemas de salud, como la cojera, también es realizada mediante observaciones visuales directas, lo que resulta en un proceso lento y a menudo impreciso. Frente a esta problemática, la incorporación de nuevas tecnologías de monitoreo en tiempo real, como el uso de dispositivos IoT, imágenes satelitales y de drones, la ciencia de datos y la inteligencia artificial, pueden ofrecer una solución eficiente y automatizada que minimiza la intervención humana y mejora la precisión del manejo ganadero, permitiendo una gestión más proactiva y basada en datos (Ma et. al., 2024; Ahmad, 2023; Lozano, et al., 2023; Wang et al., 2023; Zhou et al., 2021, Soares et al., 2021; Shao et al., 2020, Yukun et al., 2019).

La utilización de estas tecnologías en las actividades ganaderas dió inicio al nuevo paradigma conocido como Ganadería de Precisión (GP), la cual engloba un conjunto de tecnologías aplicadas al monitoreo y control en tiempo real del desempeño y bienestar de los animales. Esta metodología se apoya en la recolección y procesamiento de datos en tiempo real para generar mayor eficiencia en los procesos productivos, manejo sostenible de los recursos, reducción de costos y mejora de la calidad de los productos. Además, la GP contribuye tanto al bienestar de los animales como al de las personas encargadas de su cuidado, favoreciendo el desarrollo sostenible de las empresas agropecuarias (Lozano, et al., 2023; Wang, et al., 2023; Tzanidakis et al., 2023; Berckmans, 2017).

La GP ha evolucionado considerablemente en los últimos años, impulsada por avances tecnológicos en áreas como la teledetección, el procesamiento de imágenes y el análisis de datos en tiempo real. Sin embargo, su aplicación en sistemas ganaderos extensivos, como los que predominan en el Noroeste Argentino (NOA), aún enfrenta importantes desafíos. En estos sistemas, el ganado suele dispersarse en grandes superficies, donde los animales recorren grandes distancias en busca de pastizales naturales y agua, lo que dificulta su monitoreo continuo y eficiente (Beltramini et al., 2023; Lazarte et al., 2023; Mancuso et al., 2023; Aranda et al., 2021; Beltramini et al., 2021). Además, en la ganadería extensiva, los pastizales naturales son utilizados como fuente de forraje lo que requiere monitorear y evaluar la salud y densidad de la vegetación.

Las mejoras en la resolución de imágenes satelitales, ha facilitado el monitoreo y análisis de pastizales mediante el uso de índices de vegetación. Estos índices, que combinan diferentes bandas espectrales captadas por los satélites de teledetección, permiten resaltar la cubierta vegetal en función de su respuesta espectral, minimizando la influencia de otros factores como el suelo, la iluminación y otros elementos no vegetales.

El uso de tecnologías como collares con GPS para la localización del ganado y la obtención de imágenes satelitales y de drones para analizar la vegetación y el comportamiento animal representa una innovación clave para superar estas dificultades (Aquilani et al., 2022; horn et al., 2022; Herlin et al., 2021). La implementación de estas tecnologías permitirá detectar patrones de pastoreo, evaluar la condición corporal del ganado y realizar diagnósticos tempranos de problemas de salud del animal. Además, al analizar los índices de

vegetación y la densidad de pasturas, se podrá identificar las áreas más favorables para el pastoreo, facilitando la gestión eficiente de los recursos naturales y contribuyendo a la sostenibilidad del sistema productivo. De este modo, la transformación digital en la ganadería extensiva no sólo mejorará la productividad y reducirá los costos, sino que también garantizará un uso más responsable y sostenible de los recursos.

El presente proyecto se inscribe en este marco de innovación tecnológica y busca contribuir al desarrollo de una ganadería más eficiente, sostenible y adaptada a las nuevas demandas del mercado y del entorno productivo. La colaboración interdisciplinaria entre áreas como la electrónica, la informática y las ciencias agropecuarias será fundamental para el éxito de este proyecto, que aspira a generar un impacto positivo tanto en la productividad como en la sostenibilidad del sector ganadero

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Las principales líneas de investigación que se abordan en este proyecto son:

- Internet de las Cosas (IoT) aplicado a la ganadería: Desarrollo e implementación de dispositivos IoT, como collares con GPS, para el monitoreo en tiempo real del ganado en sistemas extensivos.
- Procesamiento de imágenes satelitales y de drones para la gestión agropecuaria: Uso de teledetección y análisis de índices de vegetación (NDVI, EVI, etc.) para evaluar la calidad y disponibilidad de los pastizales.
- Ciencia de datos y aprendizaje automático para la optimización ganadera: Aplicación de técnicas de machine learning y análisis de datos para detectar patrones de comportamiento del ganado, evaluar la condición corporal y predecir riesgos de salud.
- Transformación digital y automatización en la ganadería extensiva: Integración de tecnologías 4.0 en la producción ganadera para mejorar la toma de decisiones, optimizar recursos y promover la sostenibilidad del sector.

Para la ejecución del proyecto de I+D se planificaron las siguientes 6 etapas:

1. Configuración y despliegue de collares GPS: Se diseñarán, ajustarán e instalarán collares GPS para la recolección de datos georreferenciados, implementando un sistema de almacenamiento para facilitar el monitoreo continuo del ganado.

2. Evaluación de índices de vegetación: Se investigarán y aplicarán los índices de vegetación más adecuados (NDVI, EVI, SAVI) para monitorear la salud de los pastizales en regiones semiáridas.

3. Desarrollo de una herramienta de software para el análisis y visualización de zonas de pastoreo saludables: Se desarrollará una herramienta de software que procesará imágenes satelitales y de drones, permitiendo visualizar las áreas de pastoreo y los índices de vegetación. La herramienta destacará las zonas saludables de pastizales, facilitando la toma de decisiones informadas para optimizar la gestión del ganado y el uso de los recursos forrajeros.

4. Selección e implementación de técnicas de procesamiento de imágenes: Se revisarán técnicas de vanguardia para identificar las más adecuadas para el análisis del comportamiento del ganado en áreas extensivas del NOA, implementando y testeando algoritmos en entornos controlados.

5. Desarrollo de algoritmos para el conteo automático de ganado mediante imágenes satelitales y drones: se implementarán algoritmos avanzados de visión por computadora y redes neuronales para desarrollar un sistema de conteo automático de ganado utilizando imágenes satelitales y de drones. Este desarrollo se integrará con la herramienta de software creada en la fase 4, permitiendo que ambas funcionalidades se gestionen desde un

único sistema optimizado para la producción ganadera.

6. Publicación y difusión de resultados: Los hallazgos del proyecto serán documentados en publicaciones científicas y presentados en conferencias.

3. RESULTADOS OBTENIDOS/ESPERADOS

El grupo GIIoT se encuentra enfocado en la integración de diversas tecnologías avanzadas aplicadas a la GP. Entre sus contribuciones se destacan el desarrollo de collares con GPS para el monitoreo de ganado, la implementación de redes de largo alcance LPWAN para la transmisión de datos, el uso de visión por computadora para la clasificación de ganado, y la aplicación de ciencia de datos para el análisis y visualización del desplazamiento del ganado. A partir de los resultados obtenidos de los avances mencionados, se comenzaron estudios tendientes a potenciar el sistema de monitoreo de ganado.

En el marco del proyecto, se espera generar un conjunto de conocimientos científicos y tecnológicos que se materializarán en una herramienta avanzada para la gestión no invasiva del ganado, la cual será puesta a disposición de los ganaderos para su aplicación en el sector.

Entre los resultados esperados, se pueden mencionar:

- Desarrollo de una herramienta de software integral que visualice las zonas de pastoreo más adecuadas y realice el conteo automático de animales en áreas extensivas. Esta herramienta utilizará imágenes satelitales y de drones, junto con algoritmos avanzados de procesamiento de imágenes, para identificar y analizar las zonas de pastoreo. Al integrar funcionalidades de análisis de vegetación y conteo de ganado, la herramienta proporcionará a los ganaderos mapas interactivos y datos en tiempo real sobre la disponibilidad de recursos forrajeros y la distribución del ganado. Esto facilitará una toma de decisiones informadas y estratégicas, optimizando la gestión de los recursos, mejorando la productividad y fomentando prácticas sostenibles en la ganadería extensiva.
- Establecimiento de redes de colaboración entre ingenieros electrónicos, informáticos, agrónomos y ganaderos, fomentando el intercambio de conocimientos y prácticas.
- Publicación de los hallazgos en revistas académicas y conferencias, contribuyendo al avance del conocimiento en el ámbito de la tecnología aplicada a la ganadería de precisión.

En cuanto a la transferencia de resultados del presente proyecto a terceros beneficiarios, ya se han establecido contactos iniciales con actores clave del sector ganadero de la región del NOA, como Organismos de Investigación, asociaciones ganaderas y empresas agropecuarias interesadas en la adopción de soluciones basadas en TIC para el monitoreo y gestión del ganado. Estos contactos han manifestado interés en los resultados del proyecto, lo que indica que hay un potencial de transferencia efectiva hacia el sector.

4. FORMACIÓN DE RECURSOS HUMANOS

La línea de I+D presentada cuenta con una Doctora en Ingeniería (mención Sistemas de Información), un Magíster en Ingeniería de Computadores y Redes y dos alumnos avanzados de Ingeniería en Informática. Los profesionales del equipo son docentes investigadores Categoría III y IV.

También se prevé la capacitación y formación de recursos humanos, por medio de las siguientes actividades:

- Talleres y cursos sobre el uso de dispositivos IoT, redes de comunicación LPWAN (LoRaWAN), Machine Learning y Ciencia de Datos.
- Formación en el uso de software especializado para el análisis de índices de vegetación y monitoreo de pastizales.
- Capacitación en análisis de datos ganaderos, modelado predictivo y machine

learning aplicado a la detección de patrones de comportamiento y salud del ganado.

- Acompañamiento a estudiantes de grado y posgrado en el desarrollo de proyectos de investigación alineados con el proyecto.
- Organización de encuentros con productores, charlas técnicas y transferencia de conocimientos a comunidades agropecuarias.

5. BIBLIOGRAFÍA

Ahmad, A. A. (2023). Application of machine learning in drone technology for tracking cattle movement. Indian Journal of Animal Research, 57 (12), 1717-1724.

Aquilani, C.; Confessore, A.; Bozzi, R.; Sirtori, F.; & Pugliese, C. (2022). Precision Livestock Farming technologies in pasture-based livestock systems. Animal, 16(1), 100429.

Aranda, M. D., Redolfi, J.; Araguás, G. (2023) Aplicación de visión por computadora en ganadería de precisión en Catamarca. XX Reunión de Trabajo en Procesamiento de la Información y Control - RPIC 2023.

Aranda, M.; Beltramini, P.; Cano, J.; Virragrán, L.; Moreno, J.P.; Gallina, S.; & Herrera Conegliano, O. A. (2021). IoT aplicado a la ganadería extensiva. Revista Argentina de Ingeniería. Año 9, vol. 17. ISSN 2314-0925

Beltramini, P. I., Lazarte, I. M., Aranda, M.D., Cano, J. E., Villagrán, L. D., & Gallina, S. H. (2023). Ciencia de Datos aplicada en la ganadería de precisión en Catamarca. 1° Congreso Nacional de Ingeniería Informática / Sistemas de Información (CoNaIISI).

Herrera Conegliano, O. A., Quiroga, R. E., & Blanco, L. J. (2021). *Comportamiento y distribución animal en pastoreo extensivo*. Estación Experimental Agropecuaria La Rioja, INTA. Revista "TECNOÁRIDO" 3 (4) : 45-48.

Herrera Conegliano, O. A., Quiroga, R. E., & Blanco, L. J. (2021). *Ganadería de precisión*. Estación Experimental Agropecuaria La Rioja, INTA.

Herlin, A.; Brunberg, E.; Hultgren, J.; Högberg, N.; Rydberg, A., & Skarin, A. (2021). Animal welfare implications of digital tools for monitoring and management of cattle and sheep on pasture. Animals, 11(3), 829.

Lazarte, I.; Aranda, M.D.; Gallina, S.; Cano, J.; Beltramini, P.I. (2023). Ciencia de Datos aplicadas en la ganadería de precisión en Catamarca. 11° Congreso Nacional de Ingeniería Informática / Sistemas de Información- CoNaIISI 2023.

Lozano, J.M.l; Chayer Ricardo; Traversa, M. J. (2023). Conteo automático de ganado con imágenes tomadas con drones. Tesina de grano. Facultad de Ciencias Veterinarias. UNCPBA.

Mancuso, D.; Castagnolo, G.; & Porto, S. M. (2023). Cow Behavioural Activities in Extensive Farms: Challenges of Adopting Automatic Monitoring Systems. Sensors, 23(8), 3828.

Ma, W.; Qi, X.; Sun, Y.; Gao, R.; Ding, L.; Wang, R.; Peng, C.; Zhang, J.; Wu, J.; Xu, Z.; Li, M.; Zhao, H.; Huang, S. & Li, Q. (2024). Computer Vision-Based Measurement Techniques for Livestock Body Dimension and Weight: A Review. Agriculture, 14(2), 306.

Shao, W.; Kawakami, R.; Yoshihashi, R.; You, S.; Kawase, H.; & Naemura, T. (2020). Cattle detection and counting in UAV images based on convolutional neural networks. International Journal of Remote Sensing, 41(1), 31-52.

Soares, V. H. A.; Ponti, M. A.; Gonçalves, R. A.; & Campello, R. J. (2021). Cattle counting in the wild with geolocated aerial images in large pasture areas. *Computers and Electronics in Agriculture*, 189, 106354.

Tzanidakis, C.; Tzamaloukas, O.; Simitzis, P.; & Panagakis, P. (2023). Precision livestock farming applications (PLF) for grazing animals. Agriculture, 13(2), 288.

Wang, Y.; Mücher, S.; Wang, W.; Guo, L.; & Kooistra, L. (2023). A review of three-dimensional computer vision used in precision livestock farming for cattle growth management. Computers and Electronics in Agriculture, 206, 107687.

Yukun, S.; Pengju, H.; Yujie, W.; Ziqi, C.; Yang, L.; Baisheng, D.; Runze, L.; & Yonggen, Z. (2019). Automatic monitoring system for individual dairy cows based on a deep learning framework that provides identification via body parts and estimation of body condition score. *Journal of dairy science*, 102(11), 10140-10151.

Zhou, M.; Elmore, J. A.; Samiappan, S.; Evans, K. O.; Pfeiffer, M. B.; Blackwell, B. F.; & Iglay, R. B. (2021). Improving animal monitoring using small unmanned aircraft systems (sUAS) and deep learning networks. *Sensors*, 21(17), 5697.

Aplicaciones para adultos mayores y personas con discapacidad auditiva

María José Abásolo[12], Magdalena Rosado[3], Evelyn Del Pezo Izaguirre[4], Ana Camacho[5], Telmo Silva[6], Rita Oliveira[6], César Collazos[7], Cecilia Sanz[1], Marcelo Naiouf[1], Patricia Pesado[1], Armando De Giusti[1]

[1] Instituto de Investigación en Informática LIDI (III-LIDI)
Facultad de Informática, Universidad Nacional de La Plata (UNLP), Argentina
{mjabasolo, csanz, mnaiouf,ppesado,degiusti}@lidi.info.unlp.edu.ar
[2] Comisión de Investigaciones Científicas de la Provincia de Buenos Aires
[3] Facultad de Ingeniería Industrial, Universidad de Guayaquil, Ecuador
[4] Universidad Casa Grande (UCG), Ecuador
[5] Universidad Católica de Santiago de Guayaquil (UCSG), Ecuador
[6] Universidad de Aveiro, Portugal
{tsilva,ritaoliveira}@ua.pt
[7] Universidad del Cauca, Colombia
ccollazo@unicauca.edu.co

Resumen

La línea de investigación y desarrollo presentada consiste en estudiar, desarrollar y evaluar contenidos para la TVDi y aplicaciones móviles para el mejoramiento de los pueblos latinoamericanos. Uno de los principales objetivos es la formación de recursos humanos y fortalecimiento de la investigación mediante el trabajo intergrupal entre diferentes instituciones nacionales y extranjeras.

Palabras Clave: Televisión digital, TVDi, aplicaciones móviles, Interactividad

Contexto

Se lleva a cabo el proyecto internacional "Construcción y validación de una metodología para la transformación digital de localidades de países en desarrollo para impulsar su desarrollo sostenible" financiado por el Fondo Fiduciario Pérez Guerrero de la Oficina de Naciones Unidas para la Cooperación Sur-Sur, en el cual colaboran ESPE (Ecuador), CUJAE (Cuba) y III-LIDI UNLP (Argentina). Las instituciones forman parte de la Red de Aplicaciones y Usabilidad de la Televisión Digital Interactiva (RedAUTI) coordinada por la UNLP y conformada por diversas universidades de 11 países iberoamericanos. Han desarrollado experiencias en proyectos nacionales para investigar y desarrollar la transformación digital y se persigue lograr una coordinación efectiva para integrar y compartir los resultados obtenidos en los diferentes países, aprovechando las mejores prácticas de cada uno.
Dentro de la Facultad de Informática de la UNLP, desde hace varios años el grupo de investigación se dedica al desarrollo de aplicaciones en las áreas de Realidad Aumentada (RA), Realidad Virtual (RV), Aplicaciones Móviles y TVDi, con resultados publicados en ediciones anteriores de WICC [1-8]. Actualmente los desarrollos se realizan en el marco de los proyectos: "F031 - Diseño, desarrollo y evaluación de sistemas en escenarios híbridos para áreas clave de la sociedad actual", dirigido por C. Sanz
F032 - Computación de Alto Desempeño y Distribuida: Arquitecturas, Algoritmos, Tecnologías y Aplicaciones en HPC, Fog-Edge-Cloud, Big Data, Robótica, y Tiempo Real, dirigido por M.Naiouf

Introducción

A continuación se presentan experiencias llevadas a cabo en países latinoamericano que pueden servir de base para que otros países las promuevan para el mejoramiento de la población. Las TIC han sido utilizadas con el objetivo de mejorar su autocuidado y empoderamiento asistencial de los adultos mayores para el envejecimiento activo [9-10]. La TVDi se presenta como una herramienta prometedora para llevar a cabo experiencias con aplicaciones interactivas [11-12].
Dos de las experiencias presentadas a continuación se dirigen a adultos mayores y una de ellas a niños con discapacidad auditiva.

TVDi para la ejercitación física de adultos mayores

La tesis doctoral de Rosado se centra en la promoción de la actividad física en adultos mayores mediante la TVDi. Para ello, se desarrolló

una librería de videos con rutinas de flexibilidad, fortalecimiento muscular y equilibrio, ajustadas al nivel de riesgo de caída (leve, moderado o grave), determinado mediante pruebas funcionales de equilibrio y movilidad [13-14]. Además, se realizaron mediciones antropométricas para evaluar cambios en la composición corporal y su relación con la actividad física y la respuesta emocional de los participantes durante el uso de la TVDi fue analizada a través del maniquí de autoevaluación [15-16]. Los ejercicios son accesibles de forma remota a través de una plataforma de emisión en línea mediante Plex en un Smart TV, permitiendo su uso desde cualquier lugar con conexión a internet [17]. Los resultados del estudio evidenciaron mejoras significativas en la flexibilidad, la fuerza muscular, el equilibrio y la marcha. Asimismo, se observó una reducción en el peso corporal, la circunferencia de la cintura, el índice de masa corporal, la grasa corporal y la grasa visceral, junto con un incremento en la masa muscular magra. A nivel emocional, se identificaron diferencias estadísticamente significativas en los niveles de placer, dominio y excitación, reflejando una mayor satisfacción y bienestar durante la realización de los ejercicios enfocados en la prevención de caídas [15,18].

TVDi para ejercitar la memoria de adultos mayores

La doctorando A. Camacho actualmente tiene desarrollado un prototipo de portafolio de juegos serios para TVDi destinado a ralentizar el deterioro de la memoria en adultos mayores producto del levantamiento de información en un Centro Gerontológico [19]. El prototipo llamado Cogni-TV ha pasado por la fase de evaluación de expertos para asegurar la usabilidad del diseño [20]. Actualmente se ha concluido con un estudio inicial bajo un muestreo aleatorio simple a conveniencia como enfoque adecuado en una indagación exploratoria para facilitar la recolección de datos preliminares con el fin de realizar el siguiente estudio más detallado, introduciendo otras herramientas como tablets y celulares, siendo el objetivo la experiencia de usuario para con la plataforma Cogni-TV identificando patrones de usabilidad y aceptación. Los resultados proponen a la Televisión Digital Interactiva como un apoyo para el ralentizado de la memoria [21].

Aplicación web para enseñanza de lectura labial a niños con discapacidad auditiva

Las Tecnologías de la Información y la Comunicación (TIC) facilitan el desarrollo de alternativas complementarias de comunicación entre personas sordas y oyentes, siendo esencial que se adquieran desde edades tempranas. Esto motivó el análisis de aplicaciones educativas móviles para niños sordos [22-23], con un enfoque particular en tecnologías para la enseñanza de la lectura labial [24]. La tesis doctoral de Del Pezo presenta BlipBla, una aplicación móvil que emplea estrategias lúdicas y colaborativas para enseñar lectura labial a niños sordos en niveles básicos [25]. BlipBla es un prototipo diseñado como herramienta de apoyo para docentes, utilizándose tanto en el aula como en el hogar. Esta aplicación permite gestionar diversas opciones del juego, desde la colaboración de los profesores en la creación de material educativo hasta su aplicación directa por parte de los niños. Actualmente, se encuentra en la fase de preparación para su evaluación.

Líneas de investigación y desarrollo

Las líneas de investigación y desarrollo principales son:

- Contenidos y aplicaciones interactivas para TVDi, con especial interés en asistencia a adultos mayores
- Aplicaciones para enseñanza a discapacitados auditivos
- Metodología para la Transformación Digital adaptada a las condiciones de cada uno de los países beneficiados.
- Identificación de escenarios para impactar en las esferas priorizadas por los Objetivos de Desarrollo Sostenible.

Resultados

- Se organiza anualmente, desde 2012 a la actualidad, un evento científico que permite el intercambio de las investigaciones en el ámbito iberoamericano, denominado Jornadas de difusión y capacitación de Aplicaciones y Usabilidad de la TVDi
- Se desarrolló la aplicación MarchTV para asistir el mejoramiento de las alteraciones de la marcha de adultos mayores mediante la TVDi. Se realizó un caso de estudio en un Centro Gerontológido de

Ecuador. Recientemente se finalizó la tesis doctoral de M. Rosado
- Se desarrolló la aplicación CogniTV para entrenamiento de la memoria de adultos mayores mediante la TVDi. Se realizó un caso de estudio en un Centro Gerontológido de Ecuador. Se está desarrollando la tesis doctoral de A. Camacho.
- Se desarrolló la aplicación web Blip Bla para la enseñanza de lectura labial. Se está desarrollando la tesis doctoral de E. Del Pezo y se espera comenzar pruebas con niños con discapacidad auditiva.

Formación de recursos humanos

Participan en el proyecto PGTF dos investigadores formados de la Facultad de Informática de la UNLP, en conjunto con dos investigadores formados de CUJAE (Cuba) y tres investigadores formados de ESPE (Ecuador).
En la actualidad se están desarrollando diversas tesis de Doctorado en Ciencias Informáticas en la Facultad de Informática de la UNLP en el marco de estas líneas de investigación:
- Magdalena Rosado "Televisión Digital Interactiva (TVDi) para reducir las alteraciones de la marcha en adultos mayores". Directores: M.J.Abásolo y T. Silva (finalizada)
- Evelyn Del Pezo "Modelo colaborativo y lúdico basado en Aplicaciones Móviles para apoyar la enseñanza de la lectura labial a niños sordos" Directores: M.J.Abásolo M.J. y C. Collazos (en curso)
- Ana Camacho "Televisión Digital Interactiva (TVDi) para el retraso de Demencia Tipo Alzheimer (DTA) en Adultos Mayores" Directores: M.J.Abásolo y R.Santos. (en curso)

Referencias

[1] Abásolo M.J., De Giusti A., Naiouf M., Pesado P., Rosado Alvarez M., Pina J., Kulesza R., Silva T. (2019) *Aplicaciones de la Televisión Interactiva y tecnologías afines para el mejoramiento de los pueblos latinoamericanos.* XXI Workshop de Investigadores en Ciencias de la Computación WICC 2019, Universidad Nacional de San Juan, RedUNCI, ISBN 978-987-3619-27-4 https://sedici.unlp.edu.ar/handle/10915/77041

[2] Abásolo M.J., De Giusti A., Naiouf M., Pesado P., Sanz C., Barbieri S., Boza R., Gavilanes W., Mitaritonna A., Prinsich N., Vincenzi, M., Montero F., Perales López F. (2019) *Aplicaciones de realidad virtual, realidad aumentada e interfaces multimodales.* XXI Workshop de Investigadores en Ciencias de la Computación WICC 2019, Universidad Nacional de San Juan, RedUNCI, ISBN 978-987-3619-27-4

[3] Abásolo M.J., De Giusti A., Naiouf M., Pesado P., Rosado Alvarez M., Silva T., Pina J., Kulesza R.(2020) *Aplicaciones de la Televisión Interactiva y aplicaciones móviles para el mejoramiento de los pueblos latinoamericanos.* XXII Workshop de Investigadores en Ciencias de la Computación WICC 2020, RedUNCI, ISBN 978-987-3714-82-5 https://sedici.unlp.edu.ar/handle/10915/103733

[4] Abásolo M.J., De Giusti A., Naiouf M., Pesado P., Barbieri S., Gavilanes W., Mitaritonna A., Vincenzi, M., Bria, O.; Ronchetti, F.; Montero F., Perales López F.; Springer, V. (2020) *Aplicaciones de realidad extendida y Aplicaciones Móviles.* XXII Workshop de Investigadores en Ciencias de la Computación WICC 2020, RedUNCI, ISBN 978-987-3714-82-5

[5] Abásolo M.J., Rosado Alvarez M., Silva T., Pina J., Socorro R., Kulesza R., Silva T., Lemos de Sousa Filho G., De Giusti A., Naiouf M., Pesado P. (2021) *La Televisión Digital Interactiva para el mejoramiento de los pueblos latinoamericanos.* XXIII Workshop de Investigadores en Ciencias de la Computación WICC 2021, RedUNCI, ISBN 978-987-24611-3-3; 978-987-24611-4-0 https://sedici.unlp.edu.ar/handle/10915/120112

[6] Abásolo M.J., Gavilanes W., Del Pezo E., Allisiardi T., Bouciguez M.J., Santos G., Collazos C., De Giusti A., Naiouf M., Pesado P., Sanz C. (2021) *Realidad extendida y aplicaciones móviles aplicadas a la educación y a brindar información al ciudadano.* XXIII Workshop de Investigadores en Ciencias de la Computación WICC 2021, RedUNCI, ISBN 978-987-24611-3-3; 978-987-24611-4-0

[7] Abásolo M.J., Rosado Alvarez M., Del Pezo E.m Bouciguez M.J., Gavilanes W., Puppo F., SantosSilva T., G., Collazos C., Silva T., Ronchetti F., Sanz C., De Giusti A., Naiouf M., Pesado P. (2022) *Aplicaciones Móviles, Realidad Aumentada y TVDi.* XXIV Workshop de Investigadores en Ciencias de la Computación WICC 2022, RedUNCI, ISBN 978-987-48222-3-9 https://sedici.unlp.edu.ar/handle/10915/143936

[8] Abásolo M.J., Rosado Alvarez M., Del Pezo Izaguirre E., Camacho A., Olmedo, Pina J., Sanz C., Pesado P., Naiouf M., De Giusti A. (2024) *TVDi y Aplicaciones Móviles para la transformación digital de países en desarrollo.*

Libro de Actas del XXVI Workshop de Investigadores en Ciencias de la Computación WICC 2024, EduPA, ISBN 978-987-8352-57-2
[9] Rosado, M., Abásolo, M. J., & Silva, T. (2020). *ICT Oriented to the Elderly and Their Active Aging: A Systematic Review.* In Iberoamerican Conference on Applications and Usability of Interactive TV (pp. 134-155). Springer, Cham.
[10] Rosado M., Abásolo M.J., Silva T. (2020) *Revisión de TIC Orientadas al Adulto Mayor y su Envejecimiento Activo.* En Abásolo M.J., Kulesza R. y Pina-Amargós J. (eds) Libro de aplicaciones y usabilidad de la televisión digital interactiva: jAUTI 2019. Universidad Nacional de La Plata.
[11] Rosado M., Abásolo M.J., Silva T. (2021). *Revisión de experiencias de aplicaciones interactivas para la televisión digital ecuatoriana.* En Abreu, Abásolo, Almeida, Silva. (eds) Libro de aplicaciones y usabilidad de la televisión digital interactiva: jAUTI 2020. Universidad de Aveiro ISBN 978-972-789-680-6 https://sedici.unlp.edu.ar/handle/10915/116431
[12] Rosado M., Abásolo M.J., Silva T. (2021). *Televisores inteligentes al cuidado de la salud en adultos mayores.* En Abásolo, Olmedo (eds) Libro de aplicaciones y usabilidad de la televisión digital interactiva: jAUTI 2021. EdUNLP ISBN 978-950-34-2169-7
[13] Rosado M., Abásolo M.J., Silva T. (2021). *Contenidos interactivos para TVDI destinados a reducir las alteraciones de la marcha en adultos mayores.* En Abreu, Abásolo, Almeida, Silva. (eds) Libro de aplicaciones y usabilidad de la televisión digital interactiva: jAUTI 2020. Universidad de Aveiro . ISBN 978-972-789-680-6. https://sedici.unlp.edu.ar/handle/10915/116427
[14] Rosado M., Abásolo M.J., Silva T. (2024) *Librería de ejercicios en Smart TV para adultos mayores con riesgo de caídas.* En Abásolo, De Castro, Olmedo. (eds) Libro de aplicaciones y usabilidad de la televisión digital interactiva: jAUTI 2022 ISBN: 978-84-9927-801-8. https://sedici.unlp.edu.ar/handle/10915/162899
[15] Rosado, MM *et al.* (2024). Funcionalidad física en adultos mayores a través de la intervención de la televisión digital interactiva. En: Abásolo, MJ, Febles Estrada, A., De Castro Lozano, C. (eds) Aplicaciones y usabilidad de la televisión interactiva. jAUTI 2023. Comunicaciones en informática y ciencias de la información, vol. 2140. Springer, Cham. https://doi.org/10.1007/978-3-031-70439-0_4
[16] Rosado, M., Abásolo, M. J., & Silva, T. (2024) *Experiencia con adultos mayores en el uso de la TVDI.* En Abásolo, Febles (eds) Libro de aplicaciones y usabilidad de la televisión digital interactiva: jAUTI 2023 ISBN 978-950-34-2451-3. https://sedici.unlp.edu.ar/handle/10915/174594
[17] Rosado, M., Abásolo Guerrero, M. J., & Silva, T. (2022). Televisores inteligentes al cuidado de la salud en adultos mayores. In *X Conferencia Iberoamericana de Aplicaciones y Usabilidad de la TVDI (jAUTI 2021)(Modalidad virtual, 2 y 3 de diciembre de 2021)*.
[18] Rosado Alvarez, M. M., Espinoza Burgos, A. D., Cordova Alvarez, L. R., Valle Flores, J. A., Sierra Nieto, V. H., Jurado Auria, S. A.., Almeida Pacheco, M. E., Silva, T., & Abasolo, M. J. (2025). Prevención de caídas mediante March TV en el hogar: un estudio de caso con personas mayores independientes. *Retos*, *63*, 590–597. https://doi.org/10.47197/retos.v63.111 179
[19] Ana Camacho, M.J. Abásolo, R. Oliveira (2024) Hacia un portafolio de juegos serios para la estimulación cognitiva y de la memoria de personas con Alzheimer usando la Televisión Interactiva: Indagación en un Centro Gerontológico en Abásolo, Febles, Castro jAUTI 2023 ISBN:978-950-34-2451-3 https://sedici.unlp.edu.ar/handle/10915/174594
[20] Camacho, A., Abásolo, M. J., & Oliveira, R. (2025) (en prensa). Evaluación heurística de Cogni-TV para estimulación cognitiva de personas mayores. *Revista RISTI*.
[21] Camacho, A., Abásolo, M. J., & Oliveira, R. (2025) (en prensa). Evaluación de una Plataforma de Televisión Digital Interactiva para Adultos Mayores: Un Estudio de Experiencia de Usuario. jAUTI 2024.
[22] Del Pezo Izaguirre, E., Abásolo, M. J., & Collazos, C. A. (2020) *Uso de tecnologías móviles y realidad extendida para personas sordas: Una revisión sistemática de la literatura de acceso abierto.* Proceedings of the 15th Latin American Conference on Learning Technologies, LACLO 2020.: Institute of Electrical and Electronics Engineers
[23] E. Del Pezo Izaguirre, M. J. Abásolo and C. A. Collazos (2021) *Educational Methodologies for Hearing Impaired Children Supported by Mobile Technology and Extended Reality: Systematic Analysis of Literature*. IEEE Revista Iberoamericana de Tecnologías del Aprendizaje, vol. 16, no. 4, pp. 410-418, Nov. 2021, doi:

10.1109/RITA.2021.3135202.
[24] E. Del Pezo Izaguirre, M. J. Abásolo and C. A. Collazos (2022) *Tecnología Educativa para Enseñar la*
Lectura Labial: Un Análisis Sistemático de Literatura.Teaching lip reading to the deaf through web, mobile and extended reality technologies: a systematic review of the literature. Latin American Journal of Computing.: Escuela Politecnica Nacional. 2022 vol.9 n°2. p80 - 99. issn 1390-9266.
[25] E. Del Pezo Izaguirre, M. J. Abásolo and C. A. Collazos (2024) *Mobile and collaborative application to teach lip reading to deaf children*. Campus Virtuales.: UA Journals. 2024 vol.13 n°. p25 - 33. ISSN 2255-1514.

Desarrollo de un Modelo de procesamiento y análisis digital de imágenes satelitales para detección de cambios. Estudio de caso: Detección temprana de deforestación y degradación en la Reserva biosfera Yabotí.

María Silvia Vera Laceiras[1] Dana K. Urribarri [2]

[1]Facultad de Exactas Químicas y Naturales. UNaM. Posadas, Misiones UNNE UTN.
Facultad de Cs forestales. UNaM. Eldorado, Misiones.
Instituto de Investigación diseño e innovación. Facultad de Cs.Exactas QyN. UNaM.

[2] VyGLab(UNS-CICPBA) ICIC(UNS-CONICET) Dpto. de Cs. e Ing. de la Computación
Universidad Nacional del Sur, Bahía Blanca, Bs. As.
silvia.veralaceiras@fcf.unam.edu.ar, dku@cs.uns.edu.ar,

RESUMEN

Los bosques experimentan una creciente influencia humana, alterando servicios ecosistémicos importantes.
Mapear los bosques no solo permite muchas aplicaciones científicas, como estudiar el cambio climático y modelar la biodiversidad, sino que proporciona datos objetivos sobre los bosques permitiendo a gobiernos, sociedad civil e industria privada un abordaje preciso para mejorar el manejo forestal. El diagnóstico precoz a través de diversos algoritmos de uso de imágenes satelitales también presenta restricciones, cómo las observaciones de baja calidad, que interfieren en el análisis posterior para la clasificación y la detección.
Existen varios desafíos específicos que deben abordarse, como la alta dimensionalidad de las observaciones multiespectrales (MS) e hiperespectrales (HS); las cantidades masivas de información codificada en cada observación; las características específicas de los sensores, incluida la resolución radiométrica y las condiciones ambientales que afectan negativamente las imágenes.
El desarrollo de un modelo de procesamiento y análisis digital de imágenes satelitales para detección de cambios permitiría identificar rápidamente cambios en la estructura del paisaje, permitiendo actuar ante la deforestación de espacios protegidos. Esta información serviría para el control y diseño de prácticas sustentables en las actividades forestales de la provincia de Misiones.

Palabras claves: Detección de cambios. Teledetección. Procesamiento de imágenes.

CONTEXTO

El presente trabajo se lleva a cabo entre docentes investigadores de la facultad de ciencias forestales, (UnaM), Instituto de Investigación, diseño e innovación y de la Facultad de ciencias exactas, químicas y naturales de la UNaM y del Laboratorio de Investigación y Desarrollo en Visualización y Computación Gráfica (VyGLab, UNS-CIC Prov. de Buenos Aires) del Departamento de Ciencias e Ingeniería de la Computación, de la Universidad Nacional del Sur; además, se enmarca dentro del proyecto de "Visualización comparativa: un enfoque desde la visualización de información para la comparación de datos" (PGI 24/ZN44) de la Secretaría General de Ciencia y

Tecnología, Universidad Nacional del Sur, Argentina.

1. INTRODUCCIÓN

Tener un mapeo adecuado de la cubierta arbórea se vuelve crucial a medida que crece la necesidad de cuantificar las reservas mundiales de árboles. Específicamente, en reservas como las de Misiones, el mapeo cumple una función importante de vigilancia, dado que actualmente no existen recursos suficientes para cubrir efectivamente grandes superficies de terreno. Esto resulta en intrusiones y robos que no pueden controlarse adecuadamente.

El mapeo de la cubierta arbórea también es relevante para las políticas públicas, ya que la cuantificación de las reservas de carbono se ha considerado esencial en los tratados globales sobre la liberación y el secuestro de carbono (IGBP, 1998). Para atender estas necesidades, la comunidad de teledetección promueve los beneficios de la visión sinóptica y estandarizada proporcionada por los datos satelitales (DeFries, Hansen, Townshend, Janetos y Loveland, 2000). Uno de los productos anuales del espectrorradiómetro de imágenes de resolución moderada (MODIS) son las capas de campos continuos de vegetación. Estos mapas tienen el potencial de satisfacer muchas necesidades de la comunidad científica y de la política.

En este artículo, se describe el avance logrado con respecto a la propuesta inicial (Desarrollo de un Modelo de procesamiento y análisis digital de imágenes satelitales para detección de cambios.wicc 2024) de una metodología para obtener estimaciones del porcentaje de cobertura arbórea y la detección de cambios a lo largo del tiempo. El procedimiento, basado en el análisis de series temporales, busca identificar áreas espacialmente complejas que registran cambios debido a acciones antropogénicas.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Este trabajo se enfoca en la creación de un nuevo modelo de procesamiento y análisis digital de imágenes satelitales que permita la detección de cambios Este modelo de procesamiento y análisis digital de imágenes en su aplicación al estudio de caso permitirá validar el modelo con la detección temprana de deforestación y degradación que colabore con la conservación de la reserva de biosfera Yabotí.

Pueden analizarse a través de comparar mosaicos en más períodos de tiempo y detectar cambios en la mancha boscosa observada en la tesela del bosque nativo. Se proponen dos líneas de investigación para lograr este modelo

- Análisis multivariado y analítica visual de valores obtenidos a través de series temporales de datos e imágenes proporcionados por GEE (Google Earth Engine)
- Aplicación de algoritmos de clasificación con IA para obtención del modelo de automatización de detección de cambios.

3. RESULTADOS OBTENIDOS/ESPERADOS

Los avances obtenidos fueron: Identificación, comparación y extracción de imágenes satelitales de fuentes de datos abiertas. Se realizaron diferentes operaciones de clasificación supervisadas y no supervisadas a nivel super pixel y también se integraron las imágenes para evaluar comportamientos de las mismas. Se realizaron diferentes tipos de procesamiento estudiando las imágenes de satélites Modis, Sentinel 2 Landsat, y usando nuevos repositorios cómo el del INPE (Instituto Nacional de Pesquisas Espaciales) del Ministerio de Ciencia y Tecnología de Brasil. Se utilizaron diferentes bandas combinando en el análisis diferentes resoluciones. También se evaluó

un periodo de tiempo desde el 2000 al 2024 en diferentes estaciones para detectar cambios. El estudio recorrió varias temporadas colecciones de imágenes, metadatos, calibraciones, etc Por ejemplo se procesaron 550 imágenes obtenidas de MODIS, durante el periodo de tiempo comprendido entre los años 2000 y 2024 en 4 estaciones. Si bien la idea original en este caso, de utilizar imágenes de diferentes fechas estacionales durante cada año pretende mostrar cambios en la mancha boscosa, luego del procesamiento y análisis de estas imágenes se obtuvieron gráficos de NDVI (Índice de Vegetación de Diferencia Normalizada).en diferentes estaciones durante el lapso de tiempo 2000-2024 en los que no se pudieron observar esos cambios, aunque sí cierta ciclicidad temporal y estacional durante cada año.
Se pretende continuar con otras observaciones y registros de comportamientos temporales con satélites Landsat y con otras técnicas de procesamiento y análisis. También se buscará marcar como puntos calientes los espacios detectados con cambios y clusterizar los píxeles analizando la información que se obtiene.

4. FORMACIÓN DE RECURSOS HUMANOS

Dentro de esta línea de investigación se desarrolla una tesis de posgrado, correspondiente a la carrera de Doctorado en Ciencias de la Computación de la Universidad Nacional del Sur

5. BIBLIOGRAFÍA

Aldana-Martín, J. F., García-Nieto, J., del Mar Roldán-García, M. & Aldana Montes, J. F. (2022). Semantic modelling of Earth Observation remote sensing. Expert Systems with Applications, 187, 115838. https://doi.org/10.1016/j.eswa.2021.115838

DeFries, RS, Townshend, JRG, Hansen, M 1999 Continuous fields of vegetation characteristics at the global scale at 1km resolution. Journal of Geophysical Research, 104, 16 911 16 925.https://doi.org/10.1046/j.1365-2486.2000.00296.x

Vera Laceiras, S; Urribarri, D. K.; Ritter, L. J. & Kuna, H. (2024=.Desarrollo de un Modelo de procesamiento y análisis digital de imágenes satelitales para detección de cambios. Estudio de caso: Detección temprana de deforestación y degradación enla Reserva biosfera Yabotí. Libro de Actas XXVI Workshop de Investigadores en Ciencias de la Computación.
WICC 2024

IGBP (1998). The Terrestrial Carbon Cycle: Implications for Kyoto Protocol. Science, 280, 1393-1394. https://doi.org/10.1126/science.280.5368.1393

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C O. & Townshend,J. R. G. (2013).,). High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science*, *342*(6160), 850-853. https://doi.org/10.1126/science.1244693

Lewis, S. L., Edwards, D. P. & Galbraith, D. (2015). Increasing human dominance of tropical forests. Science, 349(6250), 827-832.
https://doi.org/10.1126/science.aaa9932

Lu, X., Jiao, L., Liu, F., Yang, S., Liu, X., Feng, Z., Li, L. & Chen, P. (2022). Simple and Efficient: A Semisupervised Learning Framework for Remote Sensing Image Semantic Segmentation. IEEE Transactions on Geoscience and Remote

Sensing, 60, 1-16. . https://doi.org/10.1109/tgrs.2022.3220755

Lu, T., Hu, Y., Fu, W., Ding, K., Bai, B. & Fang, L. (2022). SCL-Net: An End-to-End Supervised Contrastive Learning Network for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1-12. Q1. https://doi.org/10.1109/tgrs.2022.3223664

Representación cartográfica de ortofotos con datasets de mosaico—ArcMap | Documentación. (2021). Recuperado 24 de diciembre de 2022, de https://desktop.arcgis.com/es/arcmap/latest/manage-data/raster-and-images/ortho-mapping-overview.htm

N. Aburaed, M. Q. Alkhatib, S. Marshall, J. Zabalza and H. Al Ahmad,(2023) A Review of Spatial Enhancement of Hyperspectral Remote Sensing Imaging Techniques, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 16, pp. 2275-2300, https://doi:org/0.1109/JSTARS.2023.3242048.

Rosa, M. R., Brancalion, P. H. S., Crouzeilles, R., Tambosi, L. R., Piffer, P. R., Lenti, F. E. B., Hirota, M., Santiami, E. & Metzger, J. P. (2021). Hidden destruction of older forests threatens Brazil's Atlantic Forest and challenges restoration programs. *Science Advances*, *7*(4). https://doi.org/10.1126/sciadv.abc4547

Sagkatakis G, Aidini A, Fotiadou K, Giannopoulos M, Pentari A, Tsakalides P. (2019) Survey of Deep-Learning Approaches for Remote Sensing Observation Enhancement. Sensors.; 19(18):3929. https://doi.org/10.3390/s19183929

Song, X. P., Hansen, M. C., Stehman, S. V., Potapov, P. V., Tyukavina, A., Vermote, E. F. & Townshend, J. R. (2018). Global land change from 1982 to 2016. *Nature*, *560*(7720), 639-643. https://doi.org/10.1038/s41586-018-0411-9

Xue, Z., Nie, X. & Zhang, M. (2022). Incremental Dictionary Learning-Driven Tensor Low-Rank and Sparse Representation for Hyperspectral Image Classification. IEEE Transactions on Geoscience and Remote Sensing, 60, 1-19. https://doi.org/10.1109/tgrs.2022.3223792

Yu, C., Gong, B., Song, M., Zhao, E. & Chang, C. I. (2022). Multiview Calibrated Prototype Learning for Few-Shot Hyperspectral Image Classification. IEEE Transactions on Geoscience and Remote Sensing, 60, 1-13. https://doi.org/10.1109/tgrs.2022.3225947

Zhang, L. & Zhang, L. (2022). Artificial Intelligence for Remote Sensing Data Analysis: A review of challenges and opportunities. IEEE Geoscience and Remote Sensing Magazine, 10(2), 270-294. ttps://doi.org/10.1109/mgrs.2022.3145854

ESTRATEGIAS AVANZADAS PARA LOGRAR EXPERIENCIAS VIRTUALES MÁS INMERSIVAS Y PERSONALIZADAS

G. Rodríguez, N. Jofré, Y. Alvarado, J. Fernandez, R. Guerrero
Laboratorio de Computación Gráfica / Dpto. Informática / FCFMyN
Universidad Nacional de San Luis
Ejército de los Andes 950, Tel: 0266 4420823, San Luis, Argentina
{gbrodriguez, npasinetti, ymalvarado, jmfer, rag}@unsl.edu.ar

RESUMEN

El metaverso representa la convergencia entre la realidad física y digital, donde la Realidad Virtual (RV) y la Realidad Aumentada (RA) permiten una interacción más inmersiva y fluida.

La calidad de esta interacción depende en gran medida de la capacidad de los dispositivos actuales para capturar y transmitir información en tiempo real. Asimismo, para garantizar experiencias inmersivas óptimas, es esencial una comunicación fluida entre los usuarios y el entorno virtual, lo que requiere herramientas avanzadas de Interacción Humano-Computadora (IHC).

A pesar de los avances, persisten desafíos significativos, que dificultan la naturalidad y precisión de la experiencia inmersiva. La integración de tecnologías avanzadas se presenta como una solución clave para superar estas limitaciones, permitiendo mayor personalización y adaptabilidad. Esto mejoraría la interacción, optimizaría la creación de entornos dinámicos y ofrecería experiencias más realistas, con aplicaciones en entretenimiento, educación, salud y diseño de entornos virtuales.

Nuestras investigaciones actuales se enfocan en incorporar estrategias innovadoras que mejoren la percepción e interacción del usuario, estimulándolo tanto cognitiva como físicamente. Con el objetivo de evidenciar científicamente la efectividad de estas intervenciones se evaluará el impacto de sistemas de RV y RA en el desarrollo de habilidades funcionales en áreas como la salud y la educación.

Palabras clave: Metaverso, Realidad Virtual, Realidad Aumentada, Interacción Humano-Computadora, Herramientas Inteligentes.

CONTEXTO

La propuesta de trabajo se lleva a cabo dentro del proyecto "*La Realidad Extendida en la Comunicación de Información para el Metaverso*". Este proyecto es desarrollado en el ámbito del Laboratorio de Computación Gráfica de la Universidad Nacional de San Luis.

1. INTRODUCCIÓN

El metaverso se ha consolidado como un espacio virtual donde los participantes pueden intercambiar información a través de diversas actividades recreativas y laborales. Este entorno, en constante evolución, dará lugar a mundos alternativos cuyo único límite será la creatividad de sus creadores y los recursos tecnológicos disponibles. Así, en un futuro próximo, las experiencias humanas podrían desarrollarse en realidades intersubjetivas paralelas, desafiando nuestra noción tradicional de realidad [1].

Para lograr una experiencia virtual inmersiva óptima en el metaverso, es fundamental que toda la información y el conocimiento se transmitan de manera fluida entre los usuarios y el entorno virtual. Esto requiere el desarrollo de herramientas de comunicación avanzadas que permitan una interacción efectiva y eficiente tanto en la dirección usuario-metaverso como metaverso-usuario [2].

En este sentido, la Interacción Humano-Computadora (IHC) se centra en el desarrollo de dispositivos que permitan aprovechar al máximo las capacidades perceptivas humanas, logrando que la interacción sea cuasi natural [3]. Esto implica la integración de tecnologías que optimicen la comunicación entre el usuario y el entorno virtual, como interfaces hápticas, seguimiento ocular, reconocimiento de gestos y estimulación multisensorial, mejorando así la inmersión y la experiencia en el metaverso [4].

A pesar de los avances en el desarrollo del metaverso, las experiencias inmersivas aún presentan limitaciones significativas. La generación de entornos dinámicos sigue siendo un desafío, ya que requiere adaptabilidad en tiempo real y una respuesta coherente a las acciones del usuario. Además, la personalización de la experiencia sigue siendo insuficiente, lo que dificulta la creación de entornos que se ajusten a las necesidades individuales de cada usuario. Otro problema clave es la interpretación eficiente de datos sensoriales y la interacción fluida con agentes virtuales, lo que afecta la naturalidad de la inmersión [5,6].

Para superar estas limitaciones, ha sido fundamental el desarrollo de modelos computacionales avanzados mediante estrategias inteligentes [7].

Particularmente, la incorporación de Inteligencia Artificial (IA) en el metaverso es crucial para superar las limitaciones actuales en áreas como la personalización de experiencias, la adaptabilidad en tiempo real y la interpretación de datos sensoriales [8].

Su integración permitiría mejorar la fluidez de la interacción, optimizar la creación de entornos dinámicos y brindar experiencias más naturales y realistas, con aplicaciones clave en el entretenimiento, la educación, la salud y el diseño de entornos virtuales.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

La combinación de algoritmos de IA con tecnologías avanzadas en computación gráfica, Realidad Virtual y Realidad Aumentada promete mejorar la experiencia del usuario, optimizar la eficiencia de las tareas, fomentar la inclusión y aumentar la seguridad [7].

En este contexto, las siguientes líneas de investigación permitirían explorar cómo la IA puede potenciar diversos aspectos del metaverso proporcionando avances significativos.

- **Optimización de Experiencias Inmersivas:** Este campo se centra en el uso de algoritmos inteligentes para personalizar y mejorar la interacción del usuario con escenarios virtuales. Los algoritmos pueden adaptarse al comportamiento y preferencias del usuario, generando contenido interactivo y asistiendo en la navegación del metaverso, como por ejemplo: modelado del comportamiento del usuario mediante *Aprendizaje por Refuerzo* [9]*;* interpretación de movimientos de manos y expresiones faciales mediante *Redes Neuronales Convolucionales* y *Visión por Computadora* [10]; optimización de la interfaz mediante *Eye-tracking* [4], creación de texturas, personajes, escenarios o diálogos mediante *Redes Generativas Adversarias y Modelos Generativos* [11], asistentes de guía mediante *Chatbots* [12]*,* generación de animaciones realistas mediante *Motion Synthesis* [13], entre otros.

- **Accesibilidad e Inclusión en Entornos Virtuales:** Los algoritmos inteligentes juegan un papel crucial en hacer que el metaverso sea más accesible para personas con diversas discapacidades o barreras lingüísticas. Algoritmos de reconocimiento de voz [14], traducción automática y conversión de texto a voz [15], identificación de objetos o lectura

de textos [16], control de interfaces mediante movimientos oculares [17], adaptación de contenidos a las necesidades específicas de cada usuario [18], respuestas personalizadas y asistencia en tiempo real [19], interpretación de movimientos de las manos [20], entre otros, pueden colaborar en la eliminación de obstáculos.

3. RESULTADOS OBTENIDOS/ESPERADOS

Las actividades realizadas hasta el momento por este grupo se han enmarcado dentro de 1 proyecto de Investigación de la UNSL, 1 proyecto de la Comunidad Europea, 4 proyectos de Desarrollo Tecnológico de la Secretaría de Políticas Universitarias y 5 proyectos de Extensión de Interés Social.

En el marco de nuestras investigaciones previas, hemos desarrollado sistemas de RV con interacción multimodal, incorporando comandos verbales y gestuales, y aplicaciones innovadoras en RV y RA. Actualmente, nuestras acciones se centran en la integración de nuevas estrategias para mejorar la percepción e interacción del usuario, estimulándolo tanto cognitiva como física-mente.

Además, estamos analizando el impacto del uso de sistemas de RV y RA en la adquisición y mejora de habilidades funcionales en procesos físicos y cognitivos, especialmente en el área de educación. Este enfoque busca detectar y evaluar la evidencia científica para determinar la efectividad de estas interven- ciones.

4. FORMACIÓN DE RECURSOS HUMANOS

Los trabajos realizados han permitido la definición de trabajos finales de carrera de la Licenciatura en Cs. de la Computación (6 finalizados), tesis de Especialización en Educación Superior (1 finalizada), tesis de Especialización en Computación Gráfica (1 en ejecución), tesis de Maestría en Cs. de la Computación (1 finalizada y 2 en ejecución) y tesis de Doctorado en Ciencias de la Computación (1 en ejecución).

Adicionalmente se ha obtenido 2 becas de iniciación a la investigación y una beca de perfeccionamiento en investigaciones otor-gadas por la Secretaría de Ciencia y Técnica de la UNSL; y una beca doctoral de CONICET.

5. BIBLIOGRAFÍA

1. Zubillaga, A. O., & Elordi, A. G. (2022). Evolución histórica y futuro del Metaverso. Revista DYNA, 97(5), 455-457.
2. Wang, H., Ning, H., Lin, Y., Wang, W., Dhelim, S., Farha, F., ... & Daneshmand, M. (2023). A survey on the metaverse: The state-of-the-art, technologies, applications, and challenges. IEEE Internet of Things Journal, 10(16), 14671-14688.
3. Pushpakumar, R., Sanjaya, K., Rathika, S., Alawadi, A. H., Makhzuna, K., Venkatesh, S., & Rajalakshmi, B. (2023). Human-computer interaction: Enhancing user experience in interactive systems. In E3S Web of Conferences (Vol. 399, p. 04037). EDP Sciences.
4. Lyu, Z. (2023). State-of-the-Art Human-Computer-Interaction in Metaverse. International Journal of Human–Computer Interaction, 40(21), 6690–6708. https://doi.org/10.1080/10447318.2023.2248833
5. Sun, N., & Botev, J. (2023). Technological immersion and delegation to virtual agents. Multimodal Technologies and Interaction, 7(11), 106.

6. Kyrlitsias, C., & Michael-Grigoriou, D. (2022). Social Interaction With Agents and Avatars in Immersive Virtual Environments: A Survey. Frontiers in Virtual Reality.
7. Meng, Zhen & She, Changyang & Zhao, Guodong & De Martini, Daniele. (2022). Sampling, Communication, and Prediction Co-Design for Synchronizing the Real-World Device and Digital Model in Metaverse. https://doi.org/10.48550/arXiv.2208.04233
8. Thien Huynh-The, Quoc-Viet Pham, Xuan-Qui Pham, Thanh Thi Nguyen, Zhu Han, and Dong-Seong Kim. (2023). Artificial intelligence for the metaverse: A survey. Eng. Appl. Artif. Intell. 117, PA. https://doi.org/10.1016/j.engappai.2022.105581
9. Chen, X., Yao, L., Wang, X., Sun, A., Zhang, W., & Sheng, Q. Z. (2021). Generative Adversarial Reward Learning for Generalized Behavior Tendency Inference. *arXiv preprint arXiv:2105.00822*.
10. Li, K., Jin, Y., Akram, M. W., Han, R., & Chen, J. (2020). Facial expression recognition with convolutional neural networks via a new face cropping and rotation strategy. The visual computer, 36, 391-404.
11. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2020). Generative adversarial networks. Communications of the ACM, 63(11), 139-144.
12. Godwin-Jones, R. (2023). Emerging spaces for language learning: AI bots, ambient intelligence, and the metaverse. Language Learning & Technology, 27(2), 6-27.
13. Peizhuo Li, Kfir Aberman, Zihan Zhang, Rana Hanocka, and Olga Sorkine-Hornung. (2022). GANimator: neural motion synthesis from a single sequence. ACM Trans. Graph. 41, 4, Article 138 (July 2022), 12 pages. https://doi.org/10.1145/3528223.3530157
14. Li, Y., Hashim, A. S., Lin, Y., Nohuddin, P. N., Venkatachalam, K., & Ahmadian, A. (2024). AI-based visual speech recognition towards realistic avatars and lip-reading applications in the metaverse. Applied Soft Computing, 164, 111906.
15. Maza, S. B. M. (2023). Talking us into the Metaverse: Towards Realistic Streaming Speech-to-Face Animation (Doctoral dissertation, Carnegie Mellon University).
16. Xu, L., Towey, D., French, A. P., & Benford, S. (2022). Connecting everyday objects with the metaverse: A unified recognition framework. In 2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC) (pp. 401-406). IEEE.
17. Medina, B., Mejía-Rodriguez, A. R., & Alba, A. (2016). Sistema de Apoyo para Pacientes con Discapacidad Motriz Mediante una Interfaz Humano-Computadora Controlada por Movimientos Oculares. In Memorias del Congreso Nacional de Ingeniería Biomédica (Vol. 3, No. 1, pp. 214-217).
18. González, C. S. (2020). Sistema tutorial inteligente para niños con discapacidades intelectuales y cognitivas. Disertación Doctoral Universidad de La Laguna.

19. Hatami, M., Qu, Q., Chen, Y., Kholidy, H., Blasch, E., & Ardiles-Cruz, E. (2024). A Survey of the Real-Time Metaverse: Challenges and Opportunities. Future Internet, 16(10), 379.
20. Tju, T. E. E., & Shalih, M. U. (2024). Hand Sign Interpretation through Virtual Reality Data Processing. Jurnal Ilmu Komputer dan Informasi, 17(2), 185-194.

Sistema Automático de Gestión de Estacionamiento Utilizando Inteligencia Artificial

Gian Cortese[1], Julián Alejandro Lange[1], Diego A. Godoy[1], Carlos Kornuta[1], Lucas Kucuk[1], Jonathan M. Schuster [1] y Enrique M. Albornoz[2]

[1]Centro de Investigación en Tecnologías de la Información y Comunicaciones (C.I.T.I.C.) Universidad Gastón Dachary
[2]Instituto de Investigación en Señales, Sistemas e Inteligencia Computacional sinc(i), UNL-CONICET, Santa Fe, Santa Fe

gian2782cortese@gmail.com , julianlange36@gmail.com , diegodoy@citic.ugd.edu.ar, c_kornuta@citic.ugd.edu.ar, lukucuk@citic.ugd.edu.ar, jonathanmschuster@gmail.com, emalbornoz@sinc.unl.edu.ar

Resumen

Este trabajo propone el desarrollo de un sistema automático de gestión de estacionamiento utilizando inteligencia artificial. El proyecto se basa en visión artificial y *deep learning*, planificando el uso de cámaras y la red neuronal YOLO (*You Only Look Once*) para detectar vehículos y plazas de estacionamiento. La metodología incluye pruebas con diferentes versiones de YOLO para identificar el modelo más adecuado, evaluando el promedio de confianza, velocidad de inferencia y porcentaje de aciertos. Se planea utilizar Python y la librería Ultralytics para desarrollar un sistema que segmente regiones de interés y clasifique espacios como disponibles u ocupados. El objetivo es crear una solución versátil, confiable y de rápido procesamiento, aplicable a diversos escenarios.

Palabras claves: Smart parking, YOLO, visión artificial.

Contexto

Este trabajo presentado está enmarcado en el proyecto de investigación denominado "Tecnologías para Desarrollos Sostenibles de Ciudades Inteligentes apoyados por machine learning", registrado actualmente en la Secretaría de Investigación y Desarrollo de la Universidad Gastón Dachary (UGD) con el número Código IP A12001/22 (extendido hasta 2026) y radicado en el Centro de Investigación en Tecnologías de la Información y Comunicaciones de dicha universidad. El mismo fue incorporado como proyecto aprobado en el llamado a presentación interna de la UGD de proyectos de investigación N°12 mediante la Resolución Rectoral (R.R). 37-A-22 y es una continuidad de los Proyectos Tecnologías Para Desarrollos Sostenibles De Ciudades Inteligentes. R.R. UGD N° 26/A/19. Diseño de Simuladores de Procesos de Desarrollo de Software Ágiles y Redes de Sensores Inalámbricos para la Industria y la Academia. R.R. UGD N° 07/A/17 y Simulación como herramienta para la mejora de los procesos de software desarrollados con metodologías ágiles utilizando dinámica de sistemas, R.R. UGD N° 18/A/14 y R.R. UGD N° 24/A/15.

Introducción

Según estudios que se encuentran en "*Keeping our Cities Moving*" [1]

realizados por Richard Harris en 19 ciudades europeas, el 9% de los conductores pierden alrededor de 96 horas al año buscando estacionamiento y esto representa una media de 15 min. por día. Para ciudades grandes como París o Londres, estos valores tienden a incrementar y logran afectar a una media del 20% de los conductores y con un promedio de 20 minutos por día en busca de estacionamiento, esto representa una pérdida de tiempo anual de 121 horas.

El principal objetivo de este trabajo es el diseño y desarrollo de un sistema de monitoreo y gestión de estacionamiento en locaciones específicamente dispuestas para tal fin. Como se puede observar en la Figura 1, cuenta con 11 plazas de estacionamientos para vehículos de tipo sedán, hatchback, SUV, pick-up y monovolúmenes, demarcados en color rojo, 16 plazas disponibles para motocicletas cuya zona de estacionamiento está demarcada en color verde y hasta 6 bicicletas en la zona color naranja.



Figura 1. Playa de estacionamiento de la universidad.

El estacionamiento se encuentra habilitado desde las 6:00 h hasta las 22:30 h de lunes a sábado, extendiéndose desde el mes de febrero hasta el mes de diciembre. Durante el ciclo lectivo es donde se prevé una ocupación masiva de la playa de estacionamiento ya que posee mayor concurrencia de docentes, directivos, empleados administrativos y alumnos. La alta demanda de estacionamiento en el campus es un problema que este trabajo pretende analizar y brindar solución, proponiendo un sistema de gestión del uso del estacionamiento en tiempo real con el fin de proveer información del estado de ocupación de este, que puede ser consultado antes de llegar al campus. Para llevarlo a cabo se utilizarán 2 (dos) cámaras, que serán instaladas en el Campus Universitario. Dichas cámaras cumplen con requisitos mínimos de resolución de imágenes Full HD, son de uso exterior, tienen calificación de resistencia al agua IP66, tienen conexión IP y cuentan con visión nocturna. A partir de los primeros análisis de la escena, se determinó la futura ubicación de las cámaras (Figura 2), la cual es propicia para cubrir todo el estacionamiento desde dos ángulos distintos, una cámara cubre en sentido sur-norte (CAM 1), a 4 metros de altura respecto al suelo y la segunda cámara en sentido oeste-este (CAM 2), a la misma altura que la cámara mencionada anteriormente, para detectar y controlar el estado de todas las plazas de estacionamiento, evitando de esta manera, eventuales oclusiones. Las plazas tienen diferentes dimensiones que se detallan a continuación, respetando la distribución de tipos de vehículos por zonas (mencionado en el párrafo anterior), las plazas demarcadas en color rojo tienen en promedio 2 m de ancho por 3 m de largo, el sector verde tiene 8 m de ancho por 3 m de largo y el sector en color naranja tiene 2 m de ancho por 3 m de largo.

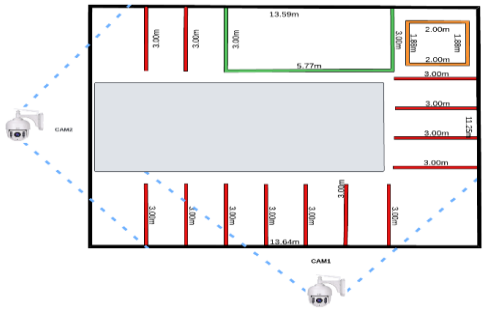


Figura 2. Plano de estacionamiento.

A continuación, se reseñan los principales conceptos y herramientas utilizadas.

Desarrollos *deep learning*: Los desarrollos de *deep learning* son aquellos que intentan imitar la forma en la que el cerebro humano procesa la información. Esto se logra mediante la utilización de redes neuronales multicapa con las que se busca que dicha red aprenda a partir de grandes cantidades de datos [2]. Los lenguajes más utilizados para cualquier desarrollo de esta índole son Python y R. Una aplicación de *deep learning* es la de clasificación de imágenes y se puede hacer uso de una Red Neuronal Convolucional (CNN) para dicha tarea [3].

PyTorch: Es una librería de aprendizaje profundo de código abierto, gratuito y de alto rendimiento basada en Python [4]. Es útil en la computación de tensores como en la implementación de redes neuronales. [5].

Open CV: Es una librería de visión artificial, permite detectar movimiento, reconocer objetos, delimitar espacios entre otras muchas funcionalidades [6].

CNN *(Convolutional Neural Networks):* Es un tipo de red neuronal con aprendizaje supervisado, que va procesando capas simulando a un cerebro humano para identificar distintas características que al final lo hacen identificar un objeto o patrón. La CNN cuenta con varias capas ocultas, que poseen una función específica y una jerarquía. En los primeros niveles las capas detectan líneas y curvas, pero al llegar a la última puede distinguir un patrón de otro [7]

YOLO *(You Only Look Once)*: Es un modelo de una red neuronal convolucional que predice los cuadros delimitadores y probabilidades de clase a partir de una imagen [8].

Línea de Investigación

En esta línea de investigación se han planteado varios objetivos. Como objetivo general se propone un prototipo que permita gestionar la playa de estacionamiento de una universidad utilizando visión artificial.

Como Objetivos específicos se plantean los siguientes: Estudiar el estado del arte de los estacionamientos inteligentes que utilizan visión artificial, analizar la detección de vehículos en delimitaciones de playas de estacionamientos, determinar los requerimientos del prototipo para detección de ocupación de playas de estacionamientos, diseñar un prototipo que permita la gestión del estacionamiento y finalmente definir un conjunto de pruebas para verificar el funcionamiento del prototipo.

Resultados

En este trabajo se detallas las etapas de la instalación de cámaras, demarcación de regiones de interés y el modelado del prototipo.

El siguiente diagrama de bloques (Figura 3) resume las etapas llevadas a cabo en el diseño y desarrollo de la solución propuesta a la problemática.

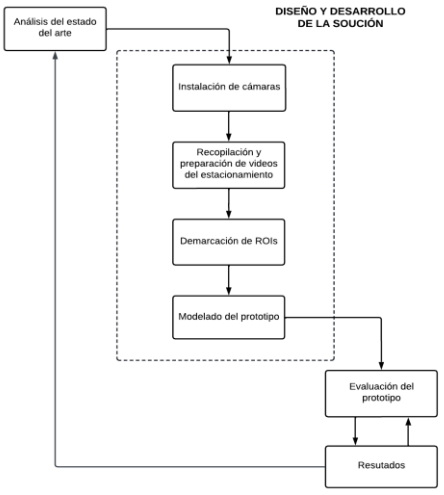


Figura 3. Diagrama de bloques del funcionamiento del sistema.

Instalación de cámaras

Se diseñó y realizó la instalación y configuración de dos cámaras de manera estratégica con la intención de evitar todas las oclusiones visuales. En la Figura 4 se observan las ubicaciones utilizadas para la colocación de las cámaras.



Figura 4. Instalación de las cámaras.

Finalizada la configuración de las cámaras se obtuvieron imágenes del estacionamiento en diferentes horarios como se observan en la Figura 5.



Figura 5. Tomas realizadas desde ambas cámaras en el estacionamiento.

Demarcación de las Regiones de Interés

La enumeración de las regiones de interés se realiza de manera incremental, denominando con números cada región donde cabe solamente un vehículo, el sector de motocicletas es el único que abarca 10 vehículos dentro del mismo ROI debido a que por el tamaño del vehículo y las diferentes posiciones en las que lo pueden estacionar, resulta factible generar un solo ROI y que la red se encargue de detectar las motocicletas se encuentran en ese espacio. En la Figura 6 se ilustran la numeración utilizada.

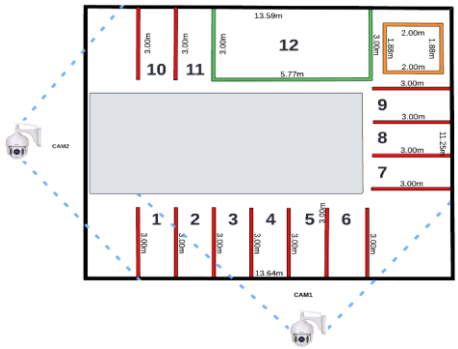


Figura 6. Numeración de cada ROI

Modelado del Prototipo

Las versiones actuales de la red neuronal convolucional YOLO se destacan por su capacidad de inferencia en videos en tiempo real, en el presente trabajo se optó por trabajar con capturas de imágenes. Esto significa que la red neuronal se activa en el momento en que se necesita, es decir, al solicitar al prototipo la disponibilidad de estacionamientos. En ese momento, se capturan las imágenes de ambas cámaras y se procesan a través de la red neuronal. El diagrama de componentes en la Figura 7 ilustran los procesos involucrados en el funcionamiento del prototipo, de forma general. Las imágenes son capturadas por las cámaras, se realiza su segmentación en ROIs y son utilizadas como entradas a la red. En función a la salida de la red y ponderaciones puntuales, se determina si una región de interés se encuentra ocupada o disponible.

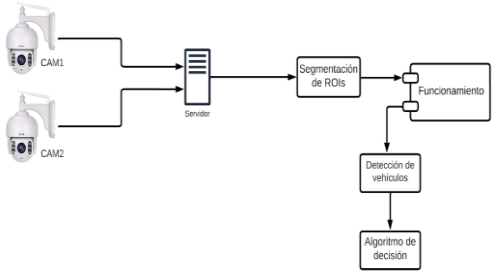


Figura 7. Procesos involucrados en el funcionamiento del prototipo.

El prototipo, cuya arquitectura se puede observar en la Figura 8, consta de una sola etapa debido a que la red neuronal elegida ya se encuentra pre entrenada. El funcionamiento del prototipo comienza con la petición mediante un dispositivo del cliente al prototipo, que se encuentra alojado en un servidor, se procede a realizar la captura de imagen de ambas cámaras, obtenidas las imágenes se determinan las regiones de interés para posteriormente analizar cada región individualmente. Las imágenes atraviesan el proceso de convolución de la red neuronal y luego se determina, mediante el algoritmo de decisión diseñado, si cada ROI se encuentra disponible u ocupado.

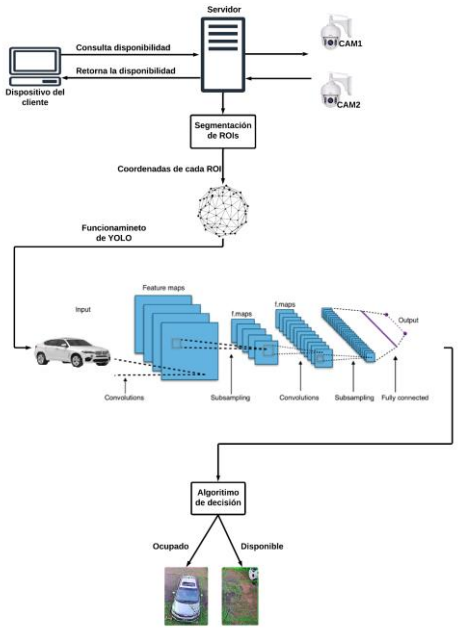


Figura 8. Arquitectura del prototipo

Formación de Recursos Humanos

El equipo encargado del desarrollo de este trabajo se encuentra formado Un Doctor en Tecnología de la Información y Comunicación y un Doctor en Ciencia y Tecnología, Un Magíster en Tecnologías de la Información y un Doctorando en Informática. Hay cuatro estudiantes en instancia de realización de trabajo final de la carrera de Ingeniería en Informática de la UGD. Además, el número de trabajos finales de carrera finalizados en este proyecto es de nueve y en curso de dos.

Bibliografía.

[1] R. Harris, «Conduent» 2017.

[2] IBM, «IBM Deep Learning,» IBM, 2022.

[3] S. Geoffrey E. Hinton, «Improving neural networks by preventing co-adaptation of feature detectors,» 03 07 2012

[4] C. Adam Paszke, «PyTorch: An Imperative Style, High-Performance Deep Learning Library,» 03 12 2019.

[5] D. Lyzhin, I. Volodchenko y L. Petrova, «Application for Recognition of Transport Using Pytorch Library,» Vladivostok, Russia, 2020.

[6] G. B. y. A. Kaeller, Learning OpenCV: Computer Vision with the OpenCV Library, Estados Unidos: O´REILLY, 2008.

[7] MathWorks, «¿Qué son las redes neuronales convolucionales?,» [En línea]. Available: https://la.mathworks.com/discovery/convolutional-neural-network-matlab.html

[8] Glenn Jocher, «Ultralytics Yolov8 Docs,» 05 12 2022.

· IS ·

Ingeniería de Software

Agrupamiento de Escenarios con Clustering no supervisado

Gladys Kaplan[1], María Inés Bertín[1], Gerardo Riera[1], Gabriel Blanco[1]

[1]Departamento de Ingeniería e Investigaciones Tecnológicas,
Universidad Nacional de La Matanza. San Justo, Buenos Aires
{gkaplan,mbertin,griera,gblanco}@unlam.edu.ar

RESUMEN

Aunque la inteligencia artificial (IA) ha revolucionado muchos aspectos del desarrollo de software y en particular, de la Ingeniería de Requisitos, todavía hay áreas donde su intervención es limitada ya que no puede reemplazar completamente el trabajo humano, como ser áreas de creatividad, negociación, intuición y la toma de decisiones humanas**.** Particularmente, en la Ingeniería de Requisitos, la IA se utiliza en la forma en que se recopilan, analizan, validan y gestionan los requisitos. Existen varias tecnologías y herramientas de IA aplicadas a esta disciplina, que incluyen Procesamiento del Lenguaje Natural (PLN), Machine Learning (ML), minería de requisitos y automatización de análisis. En el presente artículo se propone una primera aproximación al tema, extrayendo características de los escenarios futuros mediante *patrones de similitud semántica*. Posteriormente, se aplica el clustering no supervisado, en este caso *K-Means*, una técnica de *Machine Learning* que permite agrupar datos de manera automática. Este enfoque permite automatizar y mejorar la validación de requisitos, agrupar requisitos automáticamente en categorías significativas, priorizar requisitos según su relevancia en los escenarios**,** mejorar la claridad y trazabilidad, facilitar la gestión de cambios con agrupaciones dinámicas.

Palabras clave: Ingeniería de Requisitos, escenarios futuros, Machine Learning, clustering.

CONTEXTO

El presente artículo fue desarrollado en el contexto del proyecto “Impacto de la inteligencia artificial en la Ingeniería de Requisitos” como parte de la línea de investigación "Ingeniería de requisitos: Un enfoque a los negocios" que se realiza en la Universidad Nacional de La Matanza en el Departamento de Ingeniería e Investigaciones Tecnológicas.

1. INTRODUCCIÓN

El impacto de la inteligencia artificial (IA) en la Ingeniería de Requisitos (IR) [1][2][3] muestra avances significativos en la automatización y optimización en diversas etapas del proceso. La IA está revolucionando la forma en que se capturan, analizan y gestionan los requisitos, abordando problemas tradicionales de esta disciplina [4]. Ofrece herramientas para automatizar y optimizar muchas de las actividades de la IR implicando el uso de técnicas como el procesamiento de lenguaje natural (NLP), el aprendizaje automático (Machine Learning) y los sistemas basados en reglas para realizar tareas que tradicionalmente dependen de la intervención humana, como la recolección y el análisis de requisitos. Esto facilita el trabajo de los ingenieros de requisitos al reducir el tiempo que dedican a recopilar información y permite una identificación de requisitos más exhaustiva y consistente, especialmente en proyectos grandes o de rápida evolución.

La IA permite el *análisis automático* de los requisitos para identificar ambigüedades, inconsistencias y redundancias, utilizando algoritmos de NLP y modelos de aprendizaje profundo. Esto ayuda a asegurar que los requisitos estén bien definidos y sean precisos, eliminando problemas que de otra manera se detectarían en fases avanzadas del desarrollo. La detección temprana de ambigüedades **y**

errores en los requisitos mejora la calidad del software y reduce los retrabajos, generando un impacto positivo en los costos y el tiempo de desarrollo.

A pesar de los avances, existen desafíos técnicos y organizativos que dificultan la implementación generalizada de IA en la Ingeniería de Requisitos, como ser:

- Se necesitan grandes cantidades de datos de alta calidad para funcionar correctamente. Sin embargo, muchas organizaciones enfrentan dificultades en la estructuración y limpieza de datos, lo cual limita la precisión de los modelos.
- Los algoritmos pueden heredar sesgos de los datos de entrenamiento, afectando la objetividad y precisión en la priorización y análisis de requisitos.
- La adopción de soluciones utilizando IA suele implicar inversiones en tecnología y capacitación, lo que puede ser un obstáculo para empresas con recursos limitados.
- La IA sigue patrones predefinidos y puede no generar soluciones innovadoras o creativas. No siempre puede reemplazar el pensamiento crítico y la intuición humana en la definición de requisitos.
- Si se confía demasiado en la IA, se pueden perder habilidades analíticas y de toma de decisiones humanas. La IA debe ser una herramienta de apoyo, no un reemplazo total de los expertos en IR.

En el presente artículo se analizan algunas técnicas de IA con el objetivo de agrupar Escenarios Futuros (EF) [5] en el contexto del Proceso de Requisitos basado en Escenarios [6], el cual propone la construcción de los siguientes modelos:

- Un modelo léxico, LEL (Léxico Extendido del Lenguaje) el cual describe el léxico del contexto.
- Un modelo organizacional que describe el proceso del negocio tal como existe al momento del comenzar la IR, denominado Escenarios Actuales.
- Un modelo organizacional que describe el proceso del negocio con el sistema de software incluido, denominado Escenarios Futuros.
- Un modelo de requisitos donde se especifican los requisitos del software.

Cabe destacar que los EF tienen empotrados los requisitos del software, por lo tanto, el documento de especificación (ERS) se genera extrayendo los requisitos del software desde diferentes componentes del escenario. Esto pone de manifiesto la importancia de la validación de los EF. Este procedimiento presentó algunos inconvenientes como la falta de automatización, donde la IA tomó una vital relevancia.

En trabajos previos [7] se propuso una forma de agrupamiento manual por objetivos con resultados muy satisfactorios. En este procedimiento se agrupan los EF según una consigan específica, por ejemplo "¿Donde participa la OP?" y a partir de este objetivo se agrupa información de los EF en función a ciertos patrones léxicos. De esta manera se puede presentar la información del conjunto de escenarios de una forma diferente.

El presente trabajo mantiene la mirada acerca de la necesidad de agrupar los EF para una validación más segura, para capacitar a los desarrolladores y usuarios, como complemento de información para los programadores, etc. La inclusión de IA permite mejorar y automatizar el proceso y a la vez, detectar patrones [8] y tendencias no visibilizadas.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El uso del lenguaje natural (LN) ha sido un gran avance en la IR para garantizar la participación del usuario. Los modelos construidos son comprendidos por los interesados con una mínima exigencia de aprendizaje, esto se debe al formalismo que deben presentar para asegurar una correcta definición de los requisitos. En este tipo de modelo se encuentran los casos de uso, escenarios, etc. El presente trabajo se centra en los escenarios que presentan muchas ventajas, pero también algunos inconvenientes. Al ser modelos en LN los principales problemas están asociados a la ambigüedad y la falta de precisión. Estos problemas pueden ser mitigados con sintaxis, uso de glosarios, etc. Pero cuando se trata de la gestión de grandes

cantidades de escenarios, tiende a complicarse cada vez más. Esto ocurre cuando el contexto es muy extenso, participan muchos usuarios, cuando el contexto en estudio es complejo, etc. Ante una cantidad importante de escenarios la comprensión se vuelve más difícil de administrar y paradójicamente, termina por volverse confuso y arruina todo el esfuerzo realizado.

En esta investigación se utilizó el aprendizaje no supervisado, también conocido como Machine Learning no supervisado, el cual utiliza algoritmos para identificar patrones y tendencias entre los EF con el objetivo de agruparlos en clústeres. Estos algoritmos descubren agrupaciones de datos o patrones ocultos sin casi necesitar ninguna intervención humana. Cuando el resultado no es el esperado, se hace necesaria una mínima intervención del ingeniero de requisitos.

Hay varias formas de usar el clustering [9] [11], desde las exploraciones iniciales de un conjunto de datos hasta la monitorización de los procesos en curso. Puede ser utilizado en el análisis de datos exploratorios con un nuevo conjunto de datos para comprender las tendencias, los patrones y los valores atípicos subyacentes. Como alternativa, puede tener un conjunto de datos más grande que deba dividirse en varios conjuntos de datos o reducirse mediante la reducción de dimensionalidad. En estos casos, la agrupación en clúster puede ser un paso en el pre procesamiento. El proceso de clustering puede usar solo una característica de los datos o puede usar todas las características presentes en ellos.

3. RESULTADOS OBTENIDOS/ESPERADOS

Este proyecto se encuentra en la etapa inicial, por lo que los resultados generados son incipientes, pero muy prometedores. Para iniciar el estudio se utilizó el caso del "Sistema Homologaciones Universitarias". Este es un caso real donde se utilizó el *Proceso de Requisitos basado en Escenarios*. Se contó con los siguientes modelos: LEL, EA, EF, LEL_R y ERS, pero solo se utilizaron los EF y el LEL_R. Para realizar el proceso se tomaron todos títulos de los EF. Se los analizó *maximizando el uso del* ***LEL_R*** y reemplazando parte del título por palabras o frases definidas en el LEL_R, cuando era posible. Esto permitió mejorar la detección de *patrones de similitud semántica***.** En la Tabla 1 se muestran los EF involucrados.

El título del EF Nro. 7 fue modificado luego del análisis con el LEL_R. En principio el título era "Homologar materias de otras universidades", en el glosario aparecía como "universidades extranjeras", por lo cual se modificó como se muestra en Tabla 1.

Nro. EF	Título del escenario
1	Armar expediente de homologación.
2	Homologar materias de otra carrera.
3	Homologar evaluación de niveles de inglés.
4	Homologar materias por intercambio internacional.
5	Homologar asistencia y parciales por intercambio.
6	Homologar materias por doble diploma a alumnos propios.
7	Homologar materias por doble diploma para alumnos de universidades extranjera.
8	Homologar materias de otras universidades nacionales.
9	Confeccionar resolución de homologación.
10	Guardar registro de homologación
11	Guardar expediente de homologación

Tabla 1 – Escenarios Futuros

Para convertir estos textos en características clave se puede utilizar alguna de las siguientes herramientas de NLP: TF-IDF (Term Frequency-Inverse Document Frequency)**,** Embeddings de texto (Word2Vec, FastText, BERT, Sentence-BERT), Extracción de entidades (Named Entity Recognition - NER), Análisis de dependencias gramaticales**.** Esta elección incide directamente en el agrupamiento final.

Para convertir los títulos de los EF (ver Tabla 1) en números se utilizó ***TF-IDF*** que es una fórmula que calcula la relevancia de una palabra o frase en un texto. El primer paso de este análisis es detectar ***patrones de similitud***

semántica. De cada título se obtuvieron palabras clave basándose en el contenido y propósito y el resultado fue el siguiente:

Patrón 1: Procesos administrativos

- Palabras clave: ["armar", "guardar", "expediente", "registro", "resolución", "confeccionar"].

Patrón 2: Homologaciones académicas genéricas

- Palabras clave: ["materias", "otra carrera", "universidades extranjeras", "evaluación", "niveles de inglés"].

Patrón 3: Homologaciones en programas académicos

- Palabras clave: ["intercambio", "doble diploma", "asistencia", "parciales", "alumnos propios"].

Luego, ***TF-IDF*** asignó a cada EF un 0 o un 1 en función al patrón analizado, y con esta información se creó Tabla 2.

Nro. EF	Patrón 1	Patrón 2	Patrón 3
1	1	0	0
2	0	1	0
3	0	1	0
4	0	1	1
5	0	0	1
6	0	1	0
7	0	1	0
8	0	1	0
9	1	0	0
10	1	0	0
11	1	0	0

Tabla 2 – Resultado de la vectorización

Una vez encontrado los patrones, se analizaron las posibles estrategias a seguir:

1. **Agrupamiento** (Clustering). Algunas herramientas son: Agrupamiento basado en similitud de texto (***TF-IDF*** + ***K-Means*** o Word Embeddings + Clustering), Agrupamiento basado en categorías predefinidas, Agrupamiento con aprendizaje supervisado.
2. **Modelado de Temas** (Topic Modeling). Algunas herramientas son: LDA (Latent Dirichlet Allocation), NMF (Non-Negative Matrix Factorization).
3. **Clasificación Supervisada**. Algunas herramientas son: Regresión Logística, Random Forest, SVM, Redes Neuronales (BERT, GPT, LSTMs)**.**

En el presente trabajo se decidió utilizar un método de clustering denominado *K-Means* [8] [10] que requiere precisar con anticipación la cantidad de clusters (K). Al aplicar *K-Means* se puede:

- Confirmar si es correcto el agrupamiento inicial (determinar correctamente el valor de K para asegurar que el agrupamiento responde a las necesidades).
- Descubrir Patrones Ocultos.
- Reducir errores por mala interpretación de los datos.
- Agregar fácilmente nuevos escenarios y volver a procesar.

Para estimar el valor de K se puede utilizar Elbow Method, Silhouette Score, Gap Statistic, etc. En algunos casos la cantidad de clusters ya vienen identificados en la solicitud de análisis, por ejemplo, en una segmentación de mercado donde se espera agrupar en "demográfico, psicográfico, geográfico y conductual". En este caso de Homologaciones, analizando los EF se definió K=3.

El paso siguiente es ejecutar el algoritmo *K-Means*. Para ello se creó un código Python. El resultado del código es la creación de los tres clusters que se pueden observar en la Tabla 3. También se podría haber trabajado con otros valores de K y probar diferentes agrupamientos, por ejemplo, con el objetivo de capacitar, validar, priorizar requisitos, reutilizar, detectar conflictos o para trazabilidad.

El valor óptimo de K se puede analizar utilizando el Diagrama de Codo (Elbow Method).

Nro. EF	Cluster
1, 9, 10, 11	0
2, 3, 7, 8	1
4, 5, 6	2

Tabla 3 – Clusters de EF con K=3

En este caso el Método del Codo ratificó que la cantidad de clusters iniciales (K=3) era correcta, ya que el modelo puede encontrar *agrupaciones que no son evidentes* a simple vista y proponer otro valor para K. Por

ejemplo, con K=2 como se muestra en la Tabla 4.

Nro. EF	Cluster	Descripción del cluster
1, 9, 10, 11	0	Procesos Administrativos
2, 3, 4, 5, 6, 7, 8	1	Homologaciones Académicas

Tabla 4 – Clusters de EF con K=2

K-Means puede ser controlado por el usuario al definir el valor de K, la métrica de distancia y la aleatoriedad del algoritmo generando de esta manera diferentes resultados.

La adopción de IA en la IR presenta muchos beneficios, sin embargo, también presenta limitaciones en la interpretación contextual, la flexibilidad y la dependencia de datos de calidad por lo que se considera necesario combinar IA con la experiencia humana para obtener mejores resultados.

En el futuro próximo se espera analizar diferentes casos y otras técnicas y métodos de IA para determinar la que mejor se adapte a cada objetivo buscado.

4. FORMACIÓN DE RECURSOS HUMANOS

La línea de investigación presentada es parte directa de la tesis doctoral del Mg. Gabriel Blanco. También es parte de la formación inicial en investigación de la Ingeniera María Inés Bertín y del Ingeniero Gerardo Agustín Riera.

5. BIBLIOGRAFÍA

[1] Ramesh, B., & Prithi, S. (2018). "Requirements Engineering and Artificial Intelligence". Journal of Systems and Software, 137, 42-54.

[2] Kassab, M. (2019). The role of artificial intelligence in requirements engineering: A review of the state of the art. *Requirements Engineering*, *24*(4), 1-19. https://doi.org/10.1007/s00766-019-00313-8

[3] Khlood Ahmad, Muneera Bano, Mohamed Abdelrazek, Chetan Arora, John Grundy (2021) "What's up with Requirements Engineering for Artificial Intelligence Systems?", IEEE 29th International Requirements Engineering Conference (RE), USA, 2021, pp. 1-12, doi: 10.1109/RE51729.2021.00008.

[4] K. Liu, S. Reddivari and K. Reddivari, "Artificial Intelligence in Software Requirements Engineering: State-of-the-Art," 2022 IEEE 23rd International Conference on Information Reuse and Integration for Data Science (IRI), USA, 2022, pp. 106-111, doi: 10.1109/IRI54793.2022.00034.

[5] Leite, J. C. S., Hadad, G. D. S., Doorn, J. H., & Kaplan, G. N. (2000), "A scenario construction process", Requirements Engineering Journal, Springer-Verlag London Ltd., Vol.5, Nº1, pp. 38-61.

[6] Leite, J., C., S., P., Doorn, J. H., Kaplan, G., Hadad, G., D., S., Ridao, M., N., (2004) "Defining System Context using Scenarios", in Perspectives on Software Requirements, Kluwer Academic Publishers, cap.8, pp. 169-199.

[7] Kaplan Gladys N., Doorn Jorge H., Santander Candela (2022) "Agrupamiento de Escenarios guiado por objetivos", WICC22

[8] Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, *31*(8), 651-666. https://doi.org/10.1016/j.patrec.2009.09.011

[9] Aggarwal, C. C., & Reddy, C. K. (2013). *Data clustering: Algorithms and applications*. CRC Press. https://doi.org/10.1201/b15320 https://doi.org/10.1016/j.jss.2017.10.013

[10] MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1(14), 281-297.

[11] Badenes-Olmedo, C., Redondo García, J.-L., & Corcho, O. (2017). "Efficient clustering from distributions over topics", ACM Proceedings of the Knowledge Capture Conference, article 17, K-CAP 2017. https://doi.org/10.1145/3148011.3148019

ANÁLISIS DE DESEMPEÑO DE CLASIFICADORES DE TEXTO PARA PREDECIR CONFIANZA EN EQUIPOS VIRTUALES DE DESARROLLO DE SOFTWARE

Sergio Zapata[1], Gustavo Sevilla[1], Facundo Gallardo[1], Estela Torres[1]y Raymundo Forradellas[2]

[1]Instituto de Informática / Facultad de Ciencias Exactas, Físicas y Naturales / Universidad Nacional de San Juan, Argentina; {szapata, gsevilla, fgallardo, etorres}@iinfo.unsj.edu.ar

[2]Laboratorio de Sistemas Inteligentes / Universidad Nacional de Cuyo, Argentina; kike@uncu.edu.ar

RESUMEN

Actualmente una de las características que merece especial atención en los equipos virtuales de desarrollo de software es la referida a los aspectos sociales que se evidencian en las interacciones textuales de sus miembros. Comprender cómo las emociones, los estados de ánimo y especialmente la confianza entre pares afectan el resultado final de las actividades técnicas es un tema abordado por la investigación.

El texto registrado en los repositorios de software respecto de las interacciones de los miembros de estos equipos virtuales permite medir y evaluar las relaciones humanas, existiendo la posibilidad de aplicar técnicas de Procesamiento del Lenguaje Natural para medir confianza en el equipo de desarrollo de software. Esta es la línea de investigación abordada por el presente grupo de investigadores en función de la cual se muestra algunos resultados preliminares en este trabajo.

Palabras clave: Equipos Virtuales, Software Forges, Aspectos Humanos y Sociales en el Desarrollo de Software, Aprendizaje Automático, Procesamiento del Lenguaje Natural

CONTEXTO

Este trabajo de investigación está enmarcado en el proyecto “Evaluación Automática de Aspectos Sociales en Equipos Virtuales de Trabajo (ASEVI)” financiado por la Universidad Nacional de San Juan, siendo la unidad ejecutora el Instituto de Informática (IdeI) de la Facultad de Ciencias Exactas, Físicas y Naturales de la misma Universidad.

El proyecto es de ejecución bianual y su comienzo fue en enero de 2023. El mismo está inserto dentro del Programa de Incentivos a Docentes-Investigadores de la Secretaría de Políticas Universitarias del Gobierno Nacional.

1. INTRODUCCIÓN

Los equipos virtuales de desarrollo de software (EVDS) son configuraciones de equipos de trabajo muy extendidas actualmente en toda organización dedicada a la producción de software. Estos EVDS presentan las mismas ventajas y desventajas que se le atribuyen a los equipos virtuales de trabajo en general [1].

Recientemente una de las características que merece especial atención en EVDS es la referida a los aspectos sociales que se evidencian en las interacciones textuales de sus miembros. Comprender cómo las emociones, los estados de ánimo y otros aspectos humanos afectan el resultado final de las actividades técnicas (por ejemplo, la calidad del software) es un tema abordado por la investigación [2-6].

Así, la confianza entre los miembros del equipo es muy importante en el buen desempeño del proceso de desarrollo de software. En [7] se sugiere también que la confianza es un componente vital en un equipo virtual. La confianza se define como un estado

que implica expectativas positivas acerca del propósito de otro con respecto a uno mismo en el marco de situaciones de riesgo [8]. Es importante para un equipo virtual la confianza entre sus miembros ya que cohesiona al equipo mientras trabaja en una tarea en común [7], [9].

Los EVDS utilizan repositorios o forjas de software para mantener información del proyecto de software accesible a todos los miembros del equipo. Los repositorios de software son espacios virtuales donde los equipos de desarrollo generan y almacenan artefactos colaborativos a partir de las actividades de un proceso de software [10], [11], [12].

Los repositorios de software además de almacenar los artefactos, la versión final y las versiones anteriores, también almacenan las interacciones textuales entre los miembros del equipo [11].

El texto registrado en los repositorios de software respecto de las interacciones de los miembros de EVDS permite medir las relaciones humanas, existiendo ahora la posibilidad de aplicar técnicas de Procesamiento del Lenguaje Natural (PLN) para medir confianza en el equipo de desarrollo de software [13]. A continuación, mencionaremos algunos de las técnicas aplicadas bajo nuestra investigación.

Debido a su simplicidad, Naive Bayes se usa ampliamente para la clasificación de textos [14]. Naive Bayes utiliza un algoritmo de aprendizaje probabilístico simple que permite inferir la clase más probable de un documento desconocido utilizando la regla de Bayes. Con Naive Bayes se obtienen buenos resultados en la práctica, a menudo comparable a métodos de aprendizaje más sofisticados [15]. Naive Bayes Gaussiano asume que los datos siguen una distribución Gaussiana.

El Perceptrón Multicapa también de aplicación en el PLN se compone de capa de entrada, capas ocultas y capa de salida las cuales están formadas por una serie de neuronas que se encargan de recibir, procesar y enviar datos hacia otras neuronas procesando la información mediante distintas funciones matemáticas [16].

También es utilizada en el PLN la Ingeniería Prompt, que es un conjunto de técnicas y métodos para diseñar, escribir y optimizar instrucciones para Grandes Modelos de Lenguaje, llamadas prompts, de modo que las respuestas del modelo sean precisas, concretas, exactas, replicables y objetivamente correctas [17]. Según la estructura, las indicaciones pueden formularse utilizando preguntas abiertas o cerradas [18].

El método Pattern-Exploiting Training (PET) es un procedimiento de entrenamiento basado en prompts, que consiste en reformular los ejemplos de entrada como frases de estilo cloze, este tipo de frases contiene un espacio en blanco o ‘máscara’ en el texto el cual debe ser completado con una palabra teniendo en cuenta el contexto del texto, dichas palabras se las denomina etiquetas.

En los artículos elaborados por nuestro grupo de investigación, en [19], [13], se presentan análisis de los resultados de la aplicación de distintos métodos de clasificación de texto a un dataset de comentarios de EVDS extraídos desde GitHub.

2. LINEAS DE INVESTIGACIÓN Y DESARROLLO

La tendencia en las organizaciones actuales a la conformación de equipos virtuales de trabajo se mantiene firme. En estos nuevos contextos virtuales las tecnologías de la información y comunicación (TICs) permiten, entre otras funcionalidades, almacenar un conjunto de datos, como interacciones entre las personas, asignaciones de tareas, revisiones de trabajos, opiniones, etc., que de alguna manera reflejan la dinámica de trabajo del equipo.

Es en este conjunto de datos es donde nuestro grupo de investigación intenta explorar, aplicando técnicas del PLN, para obtener información valiosa de los aspectos sociales (niveles de cohesión, de colaboración, de confianza, de tono positivo, etc.) involucrados en los EVDS; reconociendo que dichos aspectos sociales son críticos para el buen desempeño de los equipos de trabajo.

El objetivo del trabajo de investigación es aportar a la Ingeniería de Software evidencia de la efectividad de métodos y modelos de lenguaje de clasificación de texto para predecir confianza en EVDS, esto a partir de la aplicación de los métodos y modelos de lenguaje sobre un dataset de comentarios resultantes de las interacciones textuales, de miembros de EVDS, registradas en la plataforma GitHub. Para esto se llevaron a cabo las siguientes actividades:

2.1. Creación de un dataset

El dataset de comentarios fue creado a partir de interacciones de equipos de desarrollo de software de organizaciones latinoamericanas. Al crear el dataset se extrajeron 1.435 comentarios en español, de tres proyectos públicos de desarrollo de software registrados en GitHub, 288 de ellos fueron de proyectos de software de México, mientras que el resto fueron de Argentina. La tarea de extracción de los comentarios se llevó a cabo utilizando la API de GitHub y el lenguaje de consulta GraphQL. Los comentarios fueron extraídos de varios pull requests (abiertos, cerrados o fusionados) de los tres proyectos de software seleccionados.

2.2. Etiquetado del dataset

Al dataset se le agregó, para cada comentario, una etiqueta de 1 o 0 según si el comentario evidenciaba confianza o no. Esta tarea de etiquetado (anotación) fue realizada por miembros del equipo de investigación. Dado un comentario, el autor (anotador) debía identificar si el comentario tenía evidencia de confianza. Varios estudios han demostrado que el tono positivo, la retroalimentación, la integridad, la delegación y el intercambio de conocimientos son características de la evidencia de confianza en los equipos de trabajo virtuales. Así los comentarios que el anotador identificaba con evidencia de confianza los etiquetaba con un 1, es decir como comentarios de confianza.

2.3. Ajuste de los Métodos y Modelos de Lenguaje y su Evaluación

Los métodos Naive Bayes, Perceptrón MC y MEP se ajustaron con el dataset etiquetado creado.

En la Figura 1. se muestra el proceso experimental seguido por los autores para la implementación en Python de los métodos Naive Bayes y Perceptron MC usando la librería sklearn, la cual integra algoritmos clásicos de aprendizaje automático.

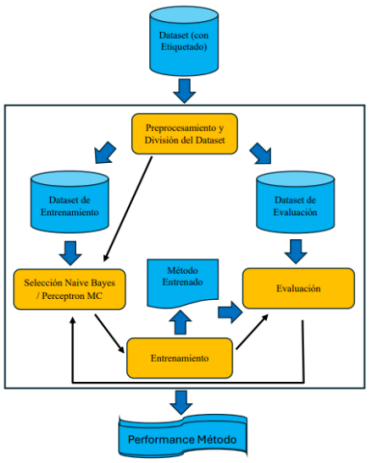


Figura 1. Implementación de los métodos Naive Bayes y Perceptrón MC para el análisis de confianza.

Así para Naive Bayes específicamente se utilizó el módulo sklearn.naive_bayes que implementa Naive Bayes con GaussianNB y el módulo sklearn.neural_network que incluye modelos basados en redes neuronales para el Perceptrón MC, en este caso se utilizó el Perceptrón MC MLPClassifier.

En la Figura 2. se observa el proceso experimental seguido por los autores para la

implementación del método PET también en Python, el cual requirió la configuración de un par patrón-verbalizador (PVP) necesario para el proceso de Fine-Tuning o ajuste fino del modelo de lenguaje, el verbalizador es una función inyectiva que mapea cada etiqueta del dataset a una palabra del vocabulario del modelo de lenguaje elegido, para el análisis de confianza se definió que la etiqueta "1" resultará en la palabra "si" del vocabulario representando a los comentarios de confianza y la etiqueta "0" en "no".

Para medir el desempeño se emplearon las métricas habitualmente usadas en tareas de clasificación de texto, estas son Precision, mide la tasa entre los comentarios correctamente identificados con confianza sobre el total de comentarios con confianza identificados; Recall mide la tasa entre los comentarios correctamente identificados con confianza sobre el total de comentarios realmente etiquetados con confianza; F1 esta métrica combina y balancea el resultado de Precision y Recall; finalmente Accuracy que mide la tasa de comentarios clasificados correctamente por el clasificador.

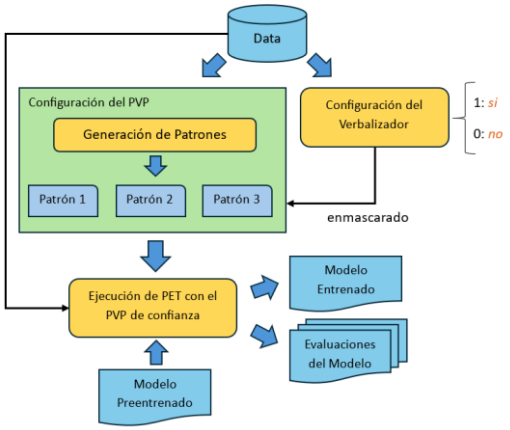


Figura 2. Implementación del método PET para el análisis de confianza.

3. RESULTADOS OBTENIDOS

Según se observa en la Tabla 1. el desempeño de los dos métodos tradicionales Naives Bayes y Perceptrón MC es moderadamente bueno, destacándose en ambos casos mejores resultados en la métrica Precision (0.70). Sin embargo, Perceptrón MC muestra un 15% de mejor desempeño en la métrica F1.

Tabla 1. Performance de los métodos y modelos de lenguaje para análisis de confianza.

Método	Modelo de Lenguaje	Precision	Recall	F1	Accuracy
Naïve Bayes		0,70	0,55	0,60	0,65
Perceptrón MC		0,70	0,69	0,69	0,68
Patter Exploiting Training-PET	BERT	0,75	0,84	0,78	0,78
	BERT-ML	0,83	0,89	0,85	0,85
	BETO	0,89	0,95	0,87	0,88
	XLM-ROBERTA	0,83	0,84	0,82	0,83
Fine-tuning	BERT-QA	0,82	0,85	0,83	0,86
	BERT-ML	0,76	0,81	0,79	0,82
	BETO	0,78	0,91	0,84	0,86
	BERT-NER	0,83	0,85	0,84	0,87

El método PET, con sus distintas variantes según modelo de lenguaje usado, está basado en los novedosos métodos de clasificación de texto que usan ingeniería de promting. Estos métodos utilizan grandes modelos de lenguaje, entrenados con grandes corpus de texto, e intentan ser muy efectivos en la definición del contexto de clasificación mediante el uso de prompts.

El método PET obtiene, en general, mejores resultados que las demás alternativas. Dentro de las distintas variantes de PET, la que usa el modelo de lenguaje BETO es la que obtiene valores más altos. Esta variante obtiene un F1 de 0.87 el cual es considerado un desempeño muy bueno.

Finalmente, los modelos de lenguaje a los que se le aplicó el método Fine-tuning, cuyos resultados son extraídos del artículo de Zapata et al. [19], obtienen buenos desempeños, especialmente la variante BETO que alcanza un F1 de 0.84.

Se observa que cuando se usa BETO como modelo de lenguaje se alcanza el mejor resultado, tanto en el método PET como en Fine-tuning, para la métrica F1. Es probable que esto se deba a que BETO fue entrenado solo con corpus de texto en español.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo lo conforman los autores del artículo, y los resultados obtenidos de la investigación forman parte de una tesis de doctorado del primer autor en el marco del Programa de Doctorado en Ciencias de la Informática que se dicta en la Facultad de Ciencias Exactas, Físicas y Naturales de la Universidad Nacional de San Juan.

5. CONCLUSIONES

La línea de investigación ejecutada realiza un aporte para instrumentar mecanismos automáticos que permitan evidenciar confianza basados en datos textuales. Para ello, utiliza y evalúa un conjunto de métodos de clasificación de texto con sus respectivas métricas de desempeño.

Se alcanzan niveles aceptables de eficacia de los clasificadores que nos permiten avizorar que es posible definir un proceso automático de medición de confianza en EVDS basados en los clasificadores de textos analizados.

Esta medición sería de gran utilidad para los líderes de proyectos de software quienes podrían tomar decisiones fundadas destinadas a la mejora del ambiente de trabajo de tales equipos. Esto es especialmente importante bajo el paradigma de desarrollo ágil, actualmente imperante, en donde las personas y sus interacciones son más importantes que los procesos y las herramientas.

6. BIBLIOGRAFÍA

[1] Törmänen, M. (2017). Virtual teams: considerations, advantages and disadvantages.

[2] Capretz, L. F., & Ahmed, F. (2010). Making sense of software development and personality types. IT professional, 12(1), 6-13.

[3] Cockburn, A., & Highsmith, J. (2001). Agile software development, the people factor. Computer, 34(11), 131-133.

[4] Graziotin, D., Wang, X., & Abrahamsson, P. (2014). Happy software developers solve problems better: psychological measurements in empirical software engineering. PeerJ, 2, e289.

[5] Novielli, N., Calefato, F., & Lanubile, F. (2014, November). Towards discovering the role of emotions in stack overflow. In Proceedings of the 6th international workshop on social software engineering (pp. 33-36).

[6] Ortu, M., Destefanis, G., Counsell, S., Swift, S., Tonelli, R., & Marchesi, M. (2017). How diverse is your team? Investigating gender and nationality diversity in GitHub teams. Journal of Software Engineering Research and Development, 5, 1-18.

[7] Greenberg, P. S., Greenberg, R. H., & Antonucci, Y. L. (2007). Creating and sustaining trust in virtual teams. Business horizons, 50(4), 325-333.

[8] Boon, S. D., & Holmes, J. G. (1991). The dynamics of interpersonal trust: Resolving uncertainty in the face of risk. Cooperation and prosocial behavior, 190, 211.

[9] Horwitz, F. M., Bravington, D., & Silvis, U. (2006). The promise of virtual teams: identifying key factors in effectiveness and failure. Journal of European Industrial Training, 30(6), 472-494.

[10] Arora, R., & Garg, A. (2018). Analysis of software repositories using process mining. In Smart Computing and Informatics: Proceedings of the First International Conference on SCI 2016, Volume 2 (pp. 637-643). Springer Singapore.

[11] Güemes-Peña, D., López-Nozal, C., Marticorena-Sánchez, R., & MaudesRaedo, J. (2018). Emerging topics in mining software repositories: Machine learning in software repositories and datasets. Progress in Artificial Intelligence, 7, 237-247.

[12] Ozbas-Caglayan, K., & Dogru, A. H. (2013, October). Software repository analysis for investigating design-code compliance. In 2013 Joint Conference of the 23rd International Workshop on Software Measurement and the 8th International Conference on Software Process and Product Measurement (pp. 231-234). IEEE.

[13] Zapata, S. G., Sevilla, G. A., Gallardo, F., Torres, E., & Forradellas, R. (2024). Predicción de confianza en equipos virtuales de desarrollo de software aplicando algoritmos de clasificación de texto. In XXX Congreso Argentino de Ciencias de la Computación (CACIC)(La Plata, 7 al 11 de octubre de 2024).

[14] Sahami, M., Dumais, S., Heckerman, D., & Horvitz, E. (1998, July). A Bayesian approach to filtering junk e-mail. In Learning for Text Categorization: Papers from the 1998 workshop (Vol. 62, pp. 98-105).

[15] Domingos, P., & Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. Machine learning, 29, 103-130.

[16] Bayona, G. A. T., & Salcedo, I. A. L. (2012). Evaluación de las redes neuronales artificiales perceptrón multicapa y fuzzy-artmap en la clasificación de imágenes satelitales. Ingeniería, 17(1), 61-72.

[17] Giray, L. (2023). Prompt engineering with ChatGPT: a guide for academic writers. Annals of biomedical engineering, 51(12), 2629-2633.

[18] Chang, E. Y. (2023, March). Prompting large language models with the socratic method. In 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC) (pp. 0351-0360). IEEE.

[19] Zapata, S., Gallardo, F., Sevilla, G., Torres, E., & Forradellas, R. (2023). Trust evaluation in virtual software development teams using BERT-based language models. Journal of Computer Science and Technology, 23(1), e04-e04.

ANÁLISIS DE MÉTODOS DE ESTIMACIÓN DE PUNTOS DE FUNCIÓN Y FUENTES DE DATOS COMPATIBLES, PARA LA AUTOMATIZACIÓN DEL PROCESO DE ESTIMAR

Nicolás Hernández, Julio Castillo, Marina Cardenas, Gonzalo Albarracín, José Victoria
Laboratorio de Investigación de Software/Dpto. Ingeniería en Sistemas de Información/Facultad Regional Córdoba/Universidad Tecnológica Nacional
{damiannicolas05, dr.jotacastillo, ing.marinacardenas, galbarracin16, josefvictoria52}@gmail.com

RESUMEN

El presente trabajo permite vislumbrar las dificultades a la hora de estimar en el desarrollo de proyectos de software, la falta de información en las etapas iniciales, cómo estos factores afectan a la gestión de proyectos, el resultado obtenido y cómo la automatización, tanto de las técnicas de recolección de información como del proceso de estimación, ayudan a mitigar los problemas mencionados.

Para garantizar el éxito de un proyecto de desarrollo de software, es importante llevar a cabo una gestión eficaz del mismo, considerando cuidadosamente las variables clave de la triple restricción: tiempo, costo y alcance. En este contexto, la estimación emerge como una herramienta fundamental en la administración de proyectos, con el propósito de anticipar con precisión el tiempo, alcance y costo asociados al desarrollo de un proyecto de software. Intentar predecir estas variables en una fase caracterizada por la escasez o ausencia de información, sumado a la elevada complejidad, dificultad y necesidad de un esfuerzo adicional inherentes a los métodos de estimación, representa un desafío significativo. A raíz de esto, se propone en el presente proyecto, el desarrollo de un sistema que permite la automatización del procedimiento de análisis de archivos, y confección de las estimaciones en las etapas de búsqueda de requerimientos y análisis del sistema. Entre los métodos de estimación que sobresalen por sus ventajas, se destacan especialmente IFPUG, NESMA y COSMI los cuales realizan el conteo de ciertas entidades para poder calcular los puntos de función, esta información puede ser obtenida de diversas fuentes de las cuales se destacan los diagramas de Lenguaje Unificado de Modelado o UML.

Palabras Clave: Automatización, Mediciones, Estimaciones, COSMIC, IFPUG y NESMA, UML, XMI, XML, JSON, SVG, Puntos de función.

CONTEXTO

El presente trabajo, representa un subproyecto, dentro del proyecto de investigación homologado por la Secretaría de Ciencia y Tecnología (SCyT) de la UTN "Modelo para el procesamiento de textos estructurados Fase 2" (cód. SIECACO0008518), el cual es desarrollado dentro del Grupo de Aprendizaje Automático, Lenguajes y Autómatas (GA^2LA), en el Laboratorio de Investigación de Software (LIS), el cual se encuentra en la Universidad Tecnológica Nacional - Facultad Regional Córdoba. El grupo antes mencionado fue creado con el objetivo de investigar las áreas de la lingüística computacional, técnicas y algoritmos de inteligencia artificial, procesamiento de lenguaje natural, de lenguajes formales, aprendizaje automático, redes neuronales, gramáticas y sintaxis [1][2][3], autómatas celulares, análisis y procesamiento de imágenes, minería de datos y desarrollo y trazabilidad de Sistemas de Información [4]. El objetivo del presente trabajo es brindar una herramienta que analice archivos de texto en formato estructurado como los son los archivos con extensión XML, XMI, JSON, y otros tipos de archivos de salida producidos por distintos tipos de software de modelado, por lo que este subproyecto queda enmarcado dentro de las áreas de investigación antes mencionadas.

1. INTRODUCCIÓN

Al iniciar un proyecto de desarrollo de software, los equipos se enfrentan al desafío de gestionar las variables de la triple restricción, lo cual implica lograr la entrega del proyecto con la funcionalidad definida, dentro del plazo acordado y respetando el presupuesto establecido. En este contexto, encontrar el equilibrio entre estas tres variables clave se convierte en una tarea esencial para el éxito y la eficiencia del proyecto.

Según la IFPUG (International Function Point User Group) el 45% de los proyectos exceden su presupuesto, 7% entregan fuera de cronograma y 56% entregan menos requerimientos [5]. Es por esta razón que, al iniciar con la administración de un nuevo proyecto, es necesario contar con herramientas que ayuden a su gestión, para lo cual se suele recurrir a la estimación.. Existen diferentes tipos de métodos de estimación, los mismos se pueden clasificar en dos categorías, la primera utiliza el conocimiento y juicio de un experto, la segunda se basa principalmente en la utilización de modelos matemáticos o métodos algorítmicos [6]. Entre los métodos del segundo grupo se pueden mencionar los métodos basados en puntos de función, los cuales miden el tamaño funcional de un sistema independientemente de la tecnología utilizada y actualmente se destacan IFPUG [7], NESMA [8]. Ambos métodos son candidatos ideales para lograr un sistema que los automatice para realizar estimaciones más precisas.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Este proyecto tiene por objetivo investigar las áreas del tratamiento y análisis de texto estructurado, el área de administración, desarrollo y control de calidad de sistemas de información. En ese sentido, la línea de investigación asociada a este trabajo se relaciona principalmente con estimaciones de software, control de calidad, desarrollo de software y análisis de texto.

A su vez, este trabajo se encuentra dentro del proyecto de "Modelo para el procesamiento de textos estructurados Fase 2" (cód. SIECACO0008518), como se mencionó anteriormente, el cual está homologado por la Secretaría de Ciencia y Tecnología de la UTN y se enmarca principalmente en el área de la Lingüística Computacional e Ingeniería de Software. Los conocimientos adquiridos en el tratamiento de texto estructurado y no estructurado permitieron llevar a cabo el análisis de los archivos de texto con los formatos XML, XMI, JSON y formatos propietarios de los distintos tipos de software de modelado de software, y los conocimientos adquiridos del área estimaciones de software en ingeniería de software, para aplicarlos en la administración de software fueron las principales líneas de investigación y desarrollo en las que se enmarca este proyecto.

3. RESULTADOS OBTENIDOS/ESPERADOS

Inicialmente, se llevó a cabo una investigación exhaustiva de diversos métodos de estimación, y en función de sus ventajas, se seleccionaron tres métodos específicos: IFPUG, NESMA y COSMIC. Una vez que se definieron estos métodos de estimación, la investigación reveló la necesidad de identificar ciertas entidades para cada uno de ellos. En consecuencia, se procedió a investigar y definir la fuente de información para obtener estas entidades. En este contexto, el Lenguaje Unificado de Modelado (UML) [10] emergió como un candidato destacado, siendo un lenguaje ampliamente utilizado actualmente durante el proceso de desarrollo de software.

Dentro del marco proporcionado por UML, se identificaron tres posibles diagramas, los cuales se seleccionaron como fuente de información esencial:

- El diagrama de casos de uso del sistema de información, que posibilita la obtención de transacciones del sistema clasificadas como entradas externas (EI), salidas externas (EO) y consultas externas (EQ).

- El diagrama de clases de análisis, que proporciona archivos de lógica interna (ILF).
- Archivos de interfaz externa (EIF), utilizados en los métodos de conteo de IFPUG y NESMA y, en el caso de COSMIC, los objetos a considerar incluyen entradas y salidas al sistema, así como lecturas y escrituras, los cuales se obtienen de los diagramas de colaboración/comunicación y secuencia.

Una vez definida la fuente de información, se procedió al estudio y selección de diversos software de modelado UML, resultando en una lista de 16 programas que cumplían con requisitos específicos, tales como soportar UML total o parcialmente, ser gratuitos o contar con una versión de prueba, y producir archivos de salida en formatos como XML, XMI, JSON o algún tipo de archivo propietario que tuviera un formato de texto estructurado. Esta información se presenta de manera detallada en la figura 1.



Figura 1. Archivo de salida XML.

Concluida la tarea de selección de software, lo siguiente fue utilizar distintos proyectos de desarrollo los cuales cumplían con la condición de ser casos similares a los de un desarrollo de software real, y con estos se realizó la transcripción de los diagramas antes mencionados a cada software seleccionado para obtener los archivos de salida, así como también la estimación en Excel de cada proyecto con los métodos de estimación, con lo que se obtuvo un banco de pruebas consistente.

Por último se desarrollaron dos componentes, un Web Service y un Cliente Web. El Web Service se compone de una clase analizadora que posee una serie de parámetros (nombre del software y tipo de archivo) que se encarga de enviar el contenido de los archivos, a otras clases analizadoras diseñadas para cada uno de los programas antes mencionados y que se corresponda con el tipo de archivo. Para cada clase analizadora se utilizaron las siguientes técnicas.

- Para los archivos XML, XMI, HTML y que presentan una estructura similar a las etiquetas utilizadas por los tres tipos mencionados, se utilizó el parser XML DOM [11], el cual se caracteriza por cargar todos los elementos en una estructura con forma de árbol, para poder crear, leer, modificar, o eliminar elementos del árbol. Esta es una técnica utilizada por los navegadores para el tratamiento de las páginas web, que serán representadas.
- Para los archivos JSON, es decir los archivos que se utilizan para el intercambio de datos de manera simple, se utilizó la librería Newtonsoft, la cual contiene una serie de herramientas para trabajar con este tipo de archivos.
- Para los archivos propietarios, se descubrió que, por lo general, los mismos tenían una estructura similar a XML o JSON, por lo que se pudieron utilizar las mismas técnicas para su análisis.

Como resultado de esto se obtuvo un Web Service que permite realizar estimaciones en base a los archivos de diagramas UML enviados, y también permite la modificación de las estimaciones realizadas cambiando la clasificación de los objetos contados y su complejidad. El mismo envía la información requerida a través del intercambio de archivos JSON a un Cliente Web que se muestra en la siguiente figura con sus respectivos métodos.

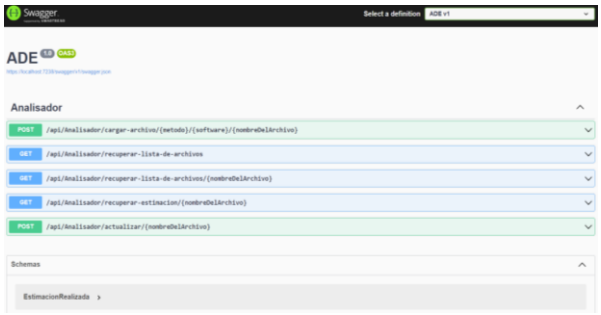


Figura 2. Web Service y sus métodos.

La principal función del Cliente Web es la de permitir la interacción con el Web Service y facilitar al usuario el uso de la herramienta, para lo cual se dispone de varias opciones que permiten realizar, consultar y modificar estimaciones e imprimirlas. La primera opción es realizar una estimación, en la cual se despliega una interfaz que permite, ingresar el nombre del archivo donde se grabará la estimación, seleccionar el método (ya sea IFPUG o NESMA), seleccionar el programa y un cuadro de diálogo para subir los archivos tal como se muestra en la imagen de la figura 3.



Figura 3. Opción para realizar una estimación.

La siguiente opción permite consultar las estimaciones realizadas representadas de manera tabular en donde se muestra, dependiendo el método seleccionado, una tabla principal con los objetos contados, su categoría, complejidad y puntos de función sin ajustar. Esta tabla puede ser modificada cambiando la complejidad y categoría de los objetos, incorporando nuevos objetos y/o eliminado lo ya existentes. La otra tabla se visualiza en el caso de que el método seleccionado sea IFPUG y se muestran los catorce factores de ajuste de los puntos de función, tal como se visualiza en la figura 4.



Figura 4.Tabla de resultados de la estimación realizada.

Al final de la pantalla se pueden ver los puntos de función en tres métodos diferentes mediante la utilización de equivalencias [12] para convertir los puntos obtenidos del método seleccionado a los otros dos métodos, por lo tanto al final contamos con resultados de IFPUG, NESMA y COSMIC como se muestra en la figura 5.

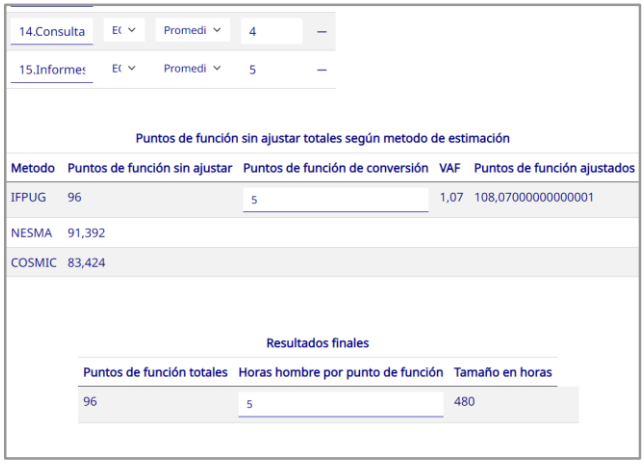


Figura 5. Resultados de IFPUG, NESMA y COSMIC.

El sistema también dispone de una opción para imprimir en formato en PDF mostrando los resultados de la estimación.

Este sistema actualmente está en etapa de prueba para lograr ajustar la precisión de las estimaciones hasta obtener el resultado esperado, que minimice las diferencias con las estimaciones realizadas en Excel.

Cabe destacar que, como trabajo futuro, se espera poder implementar opciones para estimaciones en metodologías ágiles lo cual también servirá para que estudiantes de la carrera

puedan realizar sus prácticas supervisadas y labores como becarios.

4. FORMACIÓN DE RECURSOS HUMANOS

En el proyecto de investigación dentro del cual se encuentra el trabajo descrito en este artículo participan docentes investigadores ingenieros y alumnos del departamento de Ingeniería en Sistemas de Información de la UTN FRC.

Dicho proyecto sirve como marco para la realización de la tesis de Maestría en Ingeniería en Sistemas de Información del Ing. Nicolás Hernández, cuyo director es el Dr. Julio Castillo y su co-directora, la Mg. Marina E. Cardenas. Además ha permitido que estudiantes becarios realicen diferentes tareas facilitando la incorporación de experiencia desde el punto de vista científico.

6. BIBLIOGRAFÍA

[1] Castillo J. y Cardenas M. (2010) Using Sentence Semantic Similarity Based on WordNet in Recognizing Textual Entailment. Iberamia 2010, LNCS, vol. 6433, pp. 366-375.

[2] Castillo J., Cardenas M., Curti A., Casco O., Navarro M., Hernández N. y Velazco M. (2017) Desarrollo de sistemas de análisis de texto. XIX Workshop de Investigadores en Ciencias de la Computación (WICC 2017).

[3] Cardenas M., Castillo J., Navarro M., Hernández N. y Velazco M. (2019) Herramientas para el desarrollo de sistemas

[4] Juan C Vázquez, Julio J Castillo, Leticia Constable, Marina E Cardenas. (2018). XX Workshop de Investigadores en Ciencias de la Computación (WICC 2018, Universidad Nacional del Nordeste), 2018

[5] Varajão, J., Marques, R. P., & Trigo, A. (2022). Project Management Processes – Impact on the success of information systems projects. Informática (lithuanian Academy of Sciences), 421-436. https://doi.org/10.15388/22-infor488

[6] Dapozo, D. G., Medina, M. Y., Lencina, L. B., & Petrazzini, P. G. P. (2014). Análisis comparativo de métodos de estimación basados en puntos de función para proyectos web. XX Congreso Argentino de Ciencias de la Computación (Buenos Aires, 2014). Recuperado 4 de septiembre de 2023, de http://sedici.unlp.edu.ar/handle/10915/42363

[7] Bundschuh, M., & Dekkers, C. (2008). The IFPUG Function point counting method. En Springer eBooks (pp. 323-363). https://doi.org/10.1007/978-3-540-68188-5_12

[8] NESMA. (2019) Function Point Analysis for software enhancement. Professional guide of the Nesma. https://nesma.org/ Recuperado Febrero 2023, de https://nesma.org/wp-content/uploads/2020/10/FPA-for-Software-Enhancement-v2.3-EN.pdf

[9] Consorcio Internacional de Medición de Software Común (COSMIC). (2017). El Método de Medición de Tamaño Funcional COSMIC. www.cosmic-sizing.org. Recuperado 4 de septiembre de 2023, de https://cosmic-sizing.org/publications/mmv4-0-2-diciembre2017/

[10] Koch, N., & Kraus, A. (2002, June). The expressive power of uml-based web engineering. In Second International Workshop on Web-oriented Software Technology (IWWOST02) (Vol. 16, pp. 40-41). CYTED.

[[11] Wang, F., Li, J., & Homayounfar, H. (2007). A space efficient XML DOM parser. Data & Knowledge Engineering, 60(1), 185-207.

[12] Giró, J. F. (2022). Modelos de Estimación en Ingeniería de Software (2.a ed.). (Obra original publicada 2022). ISBN: 978-987-4051-40-0.

Aplicación de Inteligencia Artificial Generativa en la Gestión de Requisitos de Software

Acosta Parra, Carlos Alberto; Barrera, María Alejandra; Bivanco, Walter Rubén; Bustos Aguiar, María Soledad; Chayle, Carolina Irene; Dahbar, Alberto; Daneri, Juan Cruz; Figueroa Vicario, Ana Carolina; Herrera, Claudia Mabel; Reynoso Chocobar, Juan Matías; Vilallonga, Gabriel Domingo

Departamento de Informática / Facultad de Tecnología y Ciencias Aplicadas
Universidad Nacional de Catamarca
Maximio Victoria Nº 55 - C.P: 4700 - San Fernando del Valle de Catamarca
Teléfono: 383- 4435112
carlosacostap@tecno.unca.edu.ar, mbarrera@tecno.unca.edu.ar, walterbivanco@gmail.com, soledadbustosaguiar@tecno.unca.edu.ar, cchayle@tecno.unca.edu.ar, albertodahbar1@gmail.com, juanjcd97@gmail.com, cfigueroa@tecno.unca.edu.ar, cm_herrera@tecno.unca.edu.ar, mattsreynoso25@gmail.com, gvilallo@tecno.unca.edu.ar

Resumen

La gestión de requisitos de software es una etapa crucial en el desarrollo de sistemas, influyendo directamente en la calidad y el éxito del producto final. Sin embargo, su ejecución presenta desafíos como la ambigüedad, inconsistencias y dificultades en la comunicación entre stakeholders. En este contexto, la inteligencia artificial generativa (IAG) ofrece oportunidades para mejorar la especificación, validación y automatización del proceso de gestión de requisitos.

Este trabajo presenta una línea de investigación orientada al diseño e implementación de un modelo basado en IAG aplicado a la gestión de requisitos, desarrollado en el marco del proyecto "Desafíos y Oportunidades de la IA Generativa". Se plantea la creación de una herramienta de software que permita evaluar su aplicabilidad en un caso de estudio real dentro de la Dirección General de Modernización de la Municipalidad de San Fernando del Valle de Catamarca. Se espera que esta solución contribuya a optimizar la precisión de los requisitos, reducir errores y mejorar la eficiencia del proceso de desarrollo de software.

Palabras Claves: Inteligencia Artificial Generativa. Gestión de Requisitos. Especificación de Software.

Contexto

Este trabajo presenta la línea de investigación centrada en la aplicación de inteligencia artificial generativa en la gestión de requisitos de software, desarrollada dentro del proyecto denominado: "Desafíos y Oportunidades de la IA Generativa: Análisis Profundo en la Transformación Digital y la Ingeniería de Software". Este proyecto es subsidiado y acreditado por la Secretaría de Investigación y Posgrado de la Universidad Nacional de Catamarca, y se desarrolla en ámbitos del Departamento de Informática de la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca, donde los integrantes del proyecto desempeñan también sus actividades como docentes y alumnos.

El proyecto se presenta como una iniciativa pionera y estratégica en el campo de la IA Generativa y se centra en la exploración de técnicas avanzadas para potenciar la transformación digital (TD) en una amplia

variedad de organizaciones, abarcando tanto el sector productivo y privado como el ámbito público. La investigación busca generar herramientas y conocimientos que puedan ser aplicados en distintos entornos, optimizando procesos y facilitando la adopción de tecnologías innovadoras.

En el segundo año de ejecución del proyecto, el enfoque se orienta a la implementación y validación del modelo de inteligencia artificial generativa aplicado a la gestión de requisitos en entornos reales. Se pretende consolidar los avances logrados en la primera etapa mediante la aplicación de las soluciones desarrolladas en organizaciones seleccionadas, permitiendo la recolección de datos y retroalimentación de usuarios. A partir de estos resultados, se buscará optimizar el modelo y generar estrategias para su adopción efectiva en contextos industriales y gubernamentales, asegurando su impacto en la mejora de la gestión de requisitos en proyectos de software.

Introducción

La gestión de requisitos es un aspecto fundamental en el desarrollo de software, ya que influye directamente en la calidad, costos y plazos de entrega de los proyectos. Sin embargo, las prácticas tradicionales de ingeniería de requisitos (RE, por sus siglas en inglés) enfrentan diversos desafíos en el manejo de sistemas de software cada vez más complejos e interconectados (Cheng et al., 2024). Según investigaciones recientes, las frecuentes modificaciones en los requisitos y la falta de comprensión de las necesidades del sistema por parte de los interesados son los principales contribuyentes a los sobrecostos en proyectos de software (Mellqvist y Mozelius, 2024).

El informe "CHAOS Report" de Standish Group (2020) reveló que apenas el 31% de los proyectos de software se completan dentro de las restricciones de tiempo y presupuesto establecidas, y solo el 46% proporcionan retornos de alto valor para los interesados (Mellqvist y Mozelius, 2024). Estos hallazgos destacan que las prácticas ineficaces de RE siguen siendo un importante cuello de botella en el desarrollo de software, lo que subraya la necesidad de enfoques innovadores para mejorar las prácticas de RE.

En este contexto, la inteligencia artificial generativa (IAG) ha emergido como una tecnología con gran potencial para transformar la gestión de requisitos, permitiendo la automatización de tareas como la elicitación, análisis, especificación, validación y gestión de requisitos funcionales y no funcionales (Cheng et al., 2024). La ingeniería de prompts se ha identificado como un factor crítico para el éxito cuando se utiliza IAG para RE, aunque los estudios recientes destacan la falta de herramientas y métodos para evaluar y crear sistemáticamente patrones de prompts efectivos para tareas específicas de RE (Mellqvist y Mozelius, 2024).

El uso de herramientas de IAG en RE presenta tanto oportunidades como desafíos significativos. Los estudios indican que mientras la IAG puede aumentar el compromiso con el aprendizaje y ayudar con la lluvia de ideas, los estudiantes y profesionales enfrentan dificultades para crear prompts efectivos y consideran que refinar los requisitos generados por IA consume mucho tiempo (Mellqvist y Mozelius, 2024). Un enfoque híbrido, que combine requisitos generados por IA y creados manualmente, ha demostrado ser más efectivo al equilibrar las ventajas de la IA con las perspectivas humanas (Mellqvist y Mozelius, 2024; Cheng et al., 2024).

Al implementar soluciones de IAG en RE, las organizaciones deben adoptar un enfoque holístico que integre marcos de validación para garantizar la consistencia y trazabilidad de los resultados, junto con estructuras de gobernanza robustas para gestionar preocupaciones éticas, de seguridad y sesgo (Cheng et al., 2024). Dado que la aplicación de GenAI en RE es un campo emergente, se requiere una investigación adicional sobre cómo integrar efectivamente estas tecnologías en la educación y práctica de la ingeniería informática.

Este trabajo explora la aplicación de IAG en la gestión de requisitos de software, con el objetivo de evaluar su impacto en la mejora de la especificación y validación de requisitos. A partir de un estudio aplicado en la Dirección de Modernización de la Municipalidad de Catamarca, se analizarán las fortalezas y limitaciones del uso de IAG en entornos reales, contribuyendo a la investigación en la aplicación de estas tecnologías en RE.

Líneas de investigación y desarrollo

Las líneas de investigación de este trabajo se centran en explorar y validar el impacto de la inteligencia artificial generativa en la gestión de requisitos de software. Para ello, se han definido las siguientes áreas clave de desarrollo:

- Revisión del estado del arte: Se llevará a cabo un análisis exhaustivo de la literatura existente sobre la aplicación de IAG en la gestión de requisitos, considerando estudios previos sobre generación de especificaciones, validación automática y mejora de la comunicación entre stakeholders.
- Diseño del modelo: Se desarrollará un modelo basado en IAG que permita la generación automática de requisitos a partir de descripciones informales, asegurando la coherencia y la eliminación de ambigüedades en el proceso de especificación. El diseño del modelo se llevará a cabo en la Dirección General de Modernización de la Municipalidad de San Fernando del Valle de Catamarca, permitiendo una retroalimentación directa de los actores involucrados en la gestión de requisitos y asegurando su alineación con las necesidades del sector público.
- Desarrollo de una herramienta: Se construirá una herramienta de software que implemente el modelo desarrollado, facilitando su aplicación en entornos reales de desarrollo de software. Esta herramienta estará diseñada para integrarse con metodologías de trabajo existentes y permitirá la automatización parcial del proceso de gestión de requisitos.
- Implementación y validación: Se aplicará el modelo y la herramienta desarrollada en un caso de estudio dentro de la Dirección General de Modernización de la Municipalidad de San Fernando del Valle de Catamarca, donde se evaluará su desempeño en entornos reales de desarrollo de software.
- Evaluación del impacto: Se medirá la eficiencia del modelo y la herramienta a través de encuestas a los usuarios involucrados en el proceso de gestión de requisitos, análisis de la calidad de los requisitos generados y comparación con enfoques tradicionales.

El objetivo de estas líneas de investigación es no solo validar el uso de IAG en la gestión de requisitos, sino también proporcionar una base para su implementación en otros contextos del desarrollo de software, optimizando la calidad y la eficiencia en el proceso de especificación.

Resultados y Objetivos

El presente trabajo tiene como objetivo general evaluar el impacto de la inteligencia artificial generativa en la gestión de requisitos de software, con un enfoque en la mejora de la calidad, precisión y eficiencia del proceso de especificación. Para ello, se establecen los siguientes objetivos específicos:

- Diseñar y desarrollar un modelo basado en IA generativa para la generación, validación y refinamiento de requisitos de software.
- Implementar una herramienta de software que permita la aplicación práctica del modelo en entornos reales.
- Aplicar el modelo y la herramienta en un caso de estudio dentro de la Dirección de Modernización de la Municipalidad de Catamarca.
- Evaluar la efectividad del modelo en la

mejora de la calidad de requisitos mediante métricas cuantitativas y cualitativas.
- Comparar los resultados obtenidos con enfoques tradicionales de gestión de requisitos, identificando ventajas y desafíos.

Se espera que la aplicación del modelo basado en IA generativa facilite la generación automática de requisitos más precisos, claros y coherentes, reduciendo errores y mejorando la comunicación entre los actores involucrados en el proceso de desarrollo de software. Además, la validación del modelo en un entorno real proporcionará evidencia empírica sobre su aplicabilidad y efectividad, permitiendo ajustes y optimizaciones para su futura implementación a mayor escala.

Asimismo, la herramienta desarrollada facilitará la integración de IA generativa en entornos de trabajo existentes, promoviendo la automatización de tareas críticas dentro de la ingeniería de requisitos y reduciendo el esfuerzo manual requerido en la generación y validación de especificaciones de software.

Formación de Recursos Humanos

La línea de investigación "Inteligencia Artificial Generativa Aplicada", enmarcada dentro de este proyecto, está compuesta por un equipo interdisciplinario de docentes, profesionales y estudiantes de grado y posgrado, quienes colaboran en la generación de conocimiento y el desarrollo de aplicaciones prácticas en el campo de la IA Generativa aplicada a la ingeniería de software.

Estructura del equipo de trabajo:

- Líderes del Proyecto: Profesores e investigadores con experiencia en IA Generativa y gestión de requisitos, responsables de la dirección del proyecto, definición de objetivos y supervisión de avances. Los mismos se han capacitado en cursos de posgrado relacionados al área del proyecto y también han dictado capacitaciones para profesionales, en ámbitos de consejos profesionales y para alumnos, en ámbitos académicos universitarios y para la comunidad en general a través del Nodo Tecnológico de Catamarca.
- Estudiantes de Posgrado: Actualmente, un estudiante de doctorado está desarrollando su tesis centrada en la aplicación de IA Generativa en la gestión de requisitos de software.
- Estudiantes de Grado: Participan en proyectos de investigación aplicada, contribuyendo al desarrollo y validación de herramientas tecnológicas relacionadas con el modelo propuesto.
- Colaboradores Externos: Profesionales del sector tecnológico y académicos de otras instituciones que aportan perspectivas y validan las soluciones desarrolladas.

Proyectos académicos en desarrollo:

Actualmente, la formación de recursos humanos dentro de este proyecto se refleja en:

- Desarrollo de una Tesis de Doctorado en Ingeniería Informática: Un integrante del equipo está desarrollando su investigación sobre "Aplicación de inteligencia artificial generativa en la gestión de requisitos de software". Este trabajo busca explorar el impacto de las técnicas de IA Generativa en la mejora de la especificación y validación de requisitos, con un enfoque en la aplicabilidad en entornos reales.
- Un Trabajo Final de grado de la carrera de Ingeniería en Informática, que están desarrollando los alumnos integrantes del proyecto, denominada: Desafíos y oportunidades de la IA Generativa: Análisis profundo en la transformación digital y la ingeniería de software.

Estos proyectos académicos no solo contribuyen al avance del conocimiento en el campo de la IA Generativa aplicada, sino que también fortalecen la capacidad del equipo para enfrentar nuevos desafíos en el ámbito de la ingeniería de

requisitos, promoviendo la generación de soluciones innovadoras y la formación de futuros expertos en la disciplina.

Bibliografía

- Cheng, H., Husen, J. H., Lu, Y., Racharak, T., Yoshioka, N., Ubayashi, N., & Washizaki, H. (2024). Generative AI for Requirements Engineering: A Systematic Literature Review. arXiv preprint arXiv:2409.06741.
- Mellqvist, N., & Mozelius, P. Exploring Student Perspectives on Generative AI in Requirements Engineering Education. In Proceedings of the International Conference on AI Research. Academic Conferences and publishing limited.
- Standish Group. (2020). Chaos Report 2020: The State of Software Development Success and Failure. Standish Group International.
- Latorre Munar, S., Jiménez Ramírez, S. A., & Ibáñez Capacho, J. S. (2024). Inteligencia Artificial Generativa para Arquitectura de Software.
- Arias Prieto, A. S., Rojas Ruiz, J. C., & Yanquen, J. C. (2023). Análisis de la inteligencia artificial generativa (IA GEN) para la gestión de riesgos en la gerencia de proyectos.
- Alto, V. (2023). Modern Generative AI with ChatGPT and OpenAI Models. Packt Publishing.
- An, J. (2019). 77 Building Blocks of Digital Transformation. Story Tree FDC.
- Caudron, J., & Van Peteghem, D. (2014). Digital transformation: A model to master digital disruption. BookBaby.
- Dervin, M. (2018). Digital Transformation from the Inside Out: Work Smarter, Increase Productivity and Create a Vibrant Culture. BookBaby.
- Farley, D. (2021). Modern Software Engineering: Doing What Works to Build Better Software Faster. Addison-Wesley Professional.
- Frankenberger, K., Mayer, H., Reiter, A., & Schmidt, M. (2020). The digital transformer's dilemma: how to energize your core business while building disruptive products and services. John Wiley & Sons.
- Gascoigne, H. (2019). The business transformation playbook: how to implement your organization’s target operating model (TOM). Hoba Tech, London.
- Goss, K. & Lakin, P. (2023). Automate it with Zapier and Generative AI. Packt Publishing.
- Greenway, A., Terrett, B., & Bracken, M. (2021). Digital transformation at scale: Why the strategy is delivery. Do Sustainability.
- Lardi, K. (2022). The Human Side of Digital Business Transformation. Wiley.
- Páez-Gabriunas, I., Sanabria, M., Gauthier-Umaña, V., Méndez-Romero, R. A., Rivera Virgüez, L., Anzola, D., ... & Saucedo Meza, G. M. (2022). Transformación digital en las organizaciones. Editorial Universidad del Rosario.
- Robertson, M. Digital Strategy: Learn To Transform Your Business for the Digital Age. Zen Mastery.
- Roth, L. J. & Johnson P. (2020). Digital transformation: An executive guide to survive and thrive in the new economy. Balboa Press.
- Sarrion, E. (2023). Exploring the Power of ChatGPT: Applications, Techniques, and Implications. Apress.
- Sommerville, I. (2020). Engineering software products (Vol. 355). London, UK: Pearson.
Vaz, N. (2021). Digital business transformation: How established companies sustain competitive advantage from now to next. John Wiley & Sons.
- Winters, T., Manshreck, T., & Wright, H. (2020). Software engineering at google: Lessons learned from programming over time. O'Reilly Media.

Aplicaciones Móviles 3D con Realidad Virtual y Realidad Aumentada

Pablo Thomas, Federico Cristina, Sebastián Dapoto, Patricia Pesado

Instituto de Investigación en Informática LIDI (III-LIDI)

Facultad de Informática – Universidad Nacional de La Plata

50 y 120 La Plata Buenos Aires

Centro Asociado a la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC)

{pthomas, fcristina, sdapoto, ppesado}@lidi.info.unlp.edu.ar

Resumen

Se presenta una línea de investigación y desarrollo, que tiene por objeto estudiar temas relacionados con aspectos de Ingeniería de Software, orientados al desarrollo de aplicaciones móviles tridimensionales (3D) sobre diversas plataformas. En particular, se pone el foco en el desarrollo de aplicaciones 3D inmersivas, a través del uso de Realidad Virtual (RV) y Realidad Aumentada (RA), y en su rendimiento gráfico y eficiencia energética. También se realiza investigación en el desarrollo de aplicaciones móviles basadas en la ubicación, haciendo uso de métodos de posicionamiento en interiores.

Palabras claves: Dispositivos Móviles – Aplicaciones 3D – Realidad Virtual – Realidad Aumentada – M-Learning – Rendimiento gráfico 3D – Eficiencia energética 3D – Posicionamiento en interiores

Contexto

Esta línea de Investigación forma parte del proyecto "*Diseño, desarrollo y evaluación de sistemas en escenarios híbridos para áreas clave de la sociedad actual: educación, ciudades inteligentes y gobernanza digital.*", en particular del subproyecto "*Ingeniería de Software para escenarios híbridos"*, del Instituto de Investigación en Informática LIDI de la Facultad de Informática, acreditado por el Ministerio de Educación de la Nación.

Existe una importante cooperación con Universidades de Argentina y se está trabajando con Universidades de Europa en proyectos financiados por el Ministerio de Ciencia y Tecnología de España y la AECID.

Se participa en iniciativas como el Programa IberoTIC de intercambio de Profesores y Alumnos de Doctorado en el área de Informática.

Introducción

En la actualidad los dispositivos móviles permiten ejecutar aplicaciones complejas y con exigentes requerimientos de hardware. Debido a esto, existen cada vez más alternativas de motores de juego que permiten desarrollar aplicaciones tridimensionales para dispositivos móviles.

La RV es una simulación interactiva por computadora en la cual se sustituye el mundo real con información sensorial que recibe el usuario. La RV permite generar entornos inmersivos donde el usuario puede interactuar con representaciones virtuales de objetos, que de otro modo sería difícil o imposible de acceder [1].

La RA es el conjunto de tecnologías que permiten que el usuario visualice el mundo real con información virtual añadida, mediante un dispositivo tecnológico. Los elementos físicos reales y tangibles se combinan con elementos virtuales. La RA es interactiva y en tiempo real [2].

El ámbito educativo debe adaptarse a los cambios y nuevas formas de aprendizaje. M-learning (mobile learning) plantea métodos modernos de apoyo al proceso de aprendizaje mediante el uso de dispositivos móviles. Las aplicaciones móviles 3D son una herramienta ideal para acercar a los alumnos [3] [4] [5].

Los sistemas de RA pueden transformar la forma de aprender y trabajar [6]. Mediante la utilización de gafas de RA y sensores inteligentes, es posible, por ejemplo, asistir a un técnico que realiza tareas de mantenimiento y deba realizar procedimientos con las manos libres.

A su vez, muchos de los potenciales usuarios de las aplicaciones 3D con RV o RA pueden no disponer de dispositivos de última generación, por lo que resulta de gran interés realizar un análisis en profundidad de los parámetros que inciden en el rendimiento gráfico de este tipo de aplicaciones.

De la misma forma, otro aspecto de relevancia es el consumo de energía de este tipo de aplicaciones, por lo que también es importante realizar un estudio que permita identificar cuáles son las características que inciden en la eficiencia energética.

Por otro lado, el posicionamiento en interiores (en inglés, indoor localization) puede ser esencial para actividades como la atención médica, la seguridad, los juegos de realidad aumentada y la asistencia a personas, entre muchos otros servicios basados en la ubicación. Aunque existen diferentes técnicas para estimar el posicionamiento en espacios interiores [7]; la triangulación WiFi se presenta como una tecnología ideal para abordar este desafío [8]. Esta técnica se basa en la medición de la intensidad de señales WiFi provenientes de puntos de acceso distribuidos en el entorno. Al analizar las diferencias de intensidad de señal entre los puntos de acceso detectados por un dispositivo móvil, es posible determinar la ubicación relativa del usuario en el espacio.

Otra técnica muy utilizada para estimar el posicionamiento en interiores es mediante el uso de los denominados *beacons*, unos pequeños dispositivos basados en tecnología Bluetooth de bajo consumo, que emiten una señal que identifica de forma única a cada dispositivo. Esta señal puede ser recibida e interpretada por un dispositivo móvil, conociendo además la distancia a la que se encuentran.

Líneas de Investigación y Desarrollo

- Metodologías y Técnicas de la Ingeniería de Software y su aplicación en el desarrollo de software para dispositivos móviles.
- Aplicaciones Móviles 3D Multiplataforma
- Mobile Learning
- Frameworks para el desarrollo de Aplicaciones Móviles 3D
- Realidad virtual y realidad aumentada en aplicaciones móviles 3D
- Rendimiento gráfico y eficiencia energética de aplicaciones móviles 3D
- Aplicaciones móviles basadas en el posicionamiento en interiores

Resultados esperados/obtenidos

Los resultados esperados/obtenidos se pueden resumir en:

- Avanzar en la capacitación continua de los miembros de la línea de investigación.
- Avanzar en el aprendizaje de frameworks o motores de juego que permiten desarrollar aplicaciones 3D multiplataforma, particularmente para dispositivos móviles [9] [10] [11] [12].
- Avanzar en el desarrollo de aplicaciones móviles 3D con realidad virtual y realidad aumentada, teniendo como finalidad enriquecer las experiencias interactivas.
- Avanzar en el estudio del rendimiento gráfico y la eficiencia energética en aplicaciones móviles 3D con realidad virtual y realidad aumentada.
- Avanzar en la investigación y desarrollo de aplicaciones móviles basadas en el posicionamiento en interiores.
- Se ha desarrollado una aplicación móvil 3D con RA denominada CGRA, que asiste a estudiantes de veterinaria en el aprendizaje de las herramientas utilizadas en las diferentes prácticas de Cirugía General Veterinaria. La aplicación brinda información, fotos y videos sobre las distintas herramientas. Además, mediante

el uso de la RA hace posible su observación en la escala original y desde cualquier punto de vista. Esto permite conocer los detalles de las herramientas y aprender el correcto uso de cada una de ellas. Figura 1.

- Se ha desarrollado una aplicación móvil 3D con RV que permite recorrer el edificio de la Facultad de Informática de la UNLP mediante el uso de gafas de RV. Además, la aplicación se integra con la API de aulas y horarios de la facultad, posibilitando informar para cada aula la materia en curso, su docente y horarios. Figura 2 y 3.
- Se ha desarrollado un prototipo de asistencia de ubicación para personas ciegas en la Facultad de Informática de la Universidad Nacional de La Plata. El prototipo permite realizar el entrenamiento preliminar necesario para contar con la base de conocimiento referencial, para que luego los usuarios finales puedan consultar su posición actual. La Figura 4 muestra el prototipo.
- Se está avanzando en el desarrollo de un prototipo que permita medir el rendimiento gráfico de las aplicaciones móviles 3D con RV.

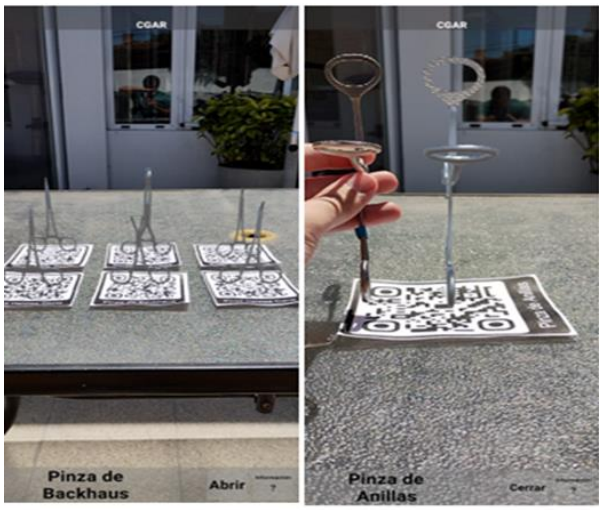


Figura 1. CGRA. Visualización de las herramientas y comparación con una herramienta real.



Figura 2. Recorrido de la Facultad de Informática de la UNLP con realidad virtual.

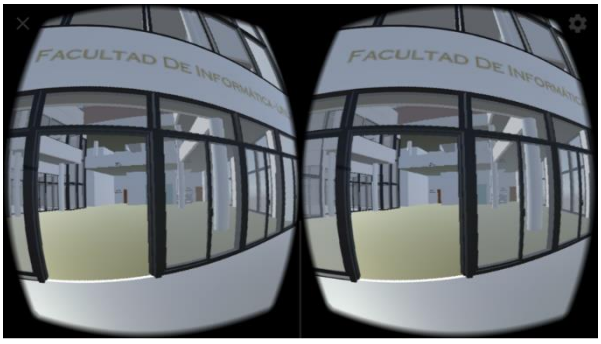


Figura 3. Recorrido de la Facultad de Informática de la UNLP. Pantalla dividida para gafas de RV.

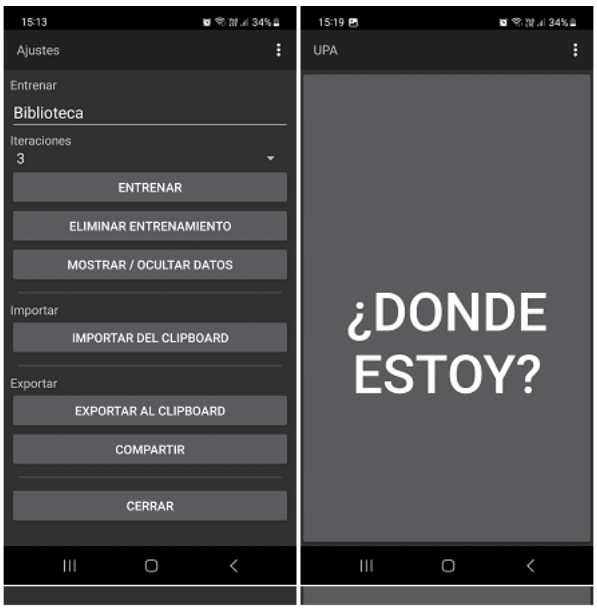


Figura 4. Prototipo de asistencia de ubicación para personas ciegas. Modo administrador (izquierda) y modo asistencia (derecha).

Formación de Recursos Humanos

Los integrantes de esta línea de investigación dirigen Prácticas Profesionales Supervisadas,

Tesinas de Grado y Tesis de Postgrado en la Facultad de Informática, y Becarios III-LIDI en temas relacionados con el proyecto. Además, participan en el dictado de asignaturas/cursos de grado y postgrado de la Facultad de Informática de la UNLP.

▪ Referencias

1. Linowes J. “Unity Virtual Reality Projects”. 2015. ISBN-13: 978-1783988556.
2. R. Silva, J.C. de Oliveira, G. Giraldi. "Introduction to augmented reality". 2003.
3. Federico Cristina, Sebastián Dapoto, Pablo Thomas, Patricia Pesado. Capítulo de Libro: “3D Mobile Prototype for Basic Algorithms Learning”. Libro: “Computer Science & Technology Series - XXI Argentine Congress Of Computer Science. Selected Papers” (300 páginas). EDULP. ISBN: 978-987-4127-00-6, páginas 239-247. Año 2016.
4. Kantel E., Tovar G., Serrano A.”Diseño de un Entorno Colaborativo Móvil para Apoyo al Aprendizaje a través de Dispositivos Móviles de Tercera Generación.” IEEE-RITA 5, no. 4 (2010): 146-151.
5. Rosa Paredes, J. Alfredo Sánchez, Liliana Rojas, Daniel Strazzulla, Ronel Martínez-Teutle. "Interacting with 3D Learning Objects". 2009 LA Web Congress. ISBN: 978-0-7695-3856-3/09.
6. X. Pan, X. Sun, H. Wang, S. Gao, N. Wang and Z. Lin, "Application of an assistant teaching system based on mobile augmented reality (AR) for course design of mechanical manufacturing process," 2017 IEEE 9th International Conference on Engineering Education (ICEED), 2017, pp. 192-196, doi: 10.1109/ICEED.2017.8251191.
7. Gu, Fuqiang & Hu, Xuke & Ramezani, Milad & Acharya, Debaditya & Khoshelham, Kourosh & Valaee, Shahrokh & Shang, Jianga. (2019). Indoor Localization Improved by Spatial Context - A Survey. ACM Computing Surveys. 52. 64:1-35. 10.1145/3322241.
8. S. Qiao, C. Cao, H. Zhou and W. Gong, "The trip to WiFi indoor localization across a decade — A systematic review," 2023 26th International Conference on Computer Supported Cooperative Work in Design (CSCWD), Rio de Janeiro, Brazil, 2023, pp. 642-647, doi: 10.1109/CSCWD57460.2023.10152700.
9. Unity 3D Homepage: https://unity3d.com/.
10. Unreal Engine Homepage. https://www.unrealengine.com/.
11. CryEngine Homepage. https://www.cryengine.com/
12. Godot Engine Homepage. https://godotengine.org

ASPECTOS DE ARQUITECTURA Y DISEÑO EN APLICACIONES DISTRIBUIDAS (WEB, MÓVILES, IOT, ETC) CON FOCO EN MICROSERVICIOS

Nicolas Battaglia, Daniel Narváez, Gustavo Rossi[1]
Centro de Altos Estudios en Tecnología Informática. Facultad de Tecnología Informática.
Universidad Abierta Interamericana. Buenos Aires. Argentina
Nicolas.Battaglia@uai.edu.ar, JoseDaniel.NarvaezF@alumnos.uai.edu.ar,
[1]También en Lifia, F. Informática, UNLP
gustavo@lifia.info.unlp.edu.ar

RESUMEN

Este proyecto busca analizar diferentes aspectos de la construcción de aplicaciones distribuidas desde una perspectiva de diseño. En este sentido se pretenden desarrollar enfoques novedosos de construcción que combinen el uso de estilos arquitecturales modernos, como microservicios, con los procesos modernos de desarrollo, como los métodos ágiles. En este trabajo enfatizamos el problema de descubrimiento de microservicios en aplicaciones "nuevas" usando métodos agiles. Presentamos brevemente la importancia de este problema dentro de la ingeniería de software, las líneas de investigación del proyecto y los resultados obtenidos desde el inicio del proyecto.

CONTEXTO

El proyecto de investigación denominado "Aspectos de Arquitectura y diseño en aplicaciones distribuidas (Web, Móviles, IoT, entre otros) con foco en Microservicios" se enmarca en la línea Ingeniería de Software y se desarrolla en conjunto entre el CAETI (Centro de Altos Estudios en Tecnología Informática) y con la Facultad de Tecnología Informática de la Universidad Abierta Interamericana (UAI).

1. INTRODUCCIÓN

1.1 Microservicios

El crecimiento de las arquitecturas basadas en microservicios (MS) se debe a sus ventajas frente a las arquitecturas monolíticas, especialmente al aprovechar la computación en la nube. Los MS facilitan la mantenibilidad y escalabilidad mediante el desarrollo, prueba y despliegue independiente de cada servicio, mejorando la flexibilidad y reduciendo la complejidad. Su modularidad permite que equipos distribuidos trabajen de forma autónoma, acelerando el desarrollo. Además, su tolerancia a fallos asegura que la falla de un servicio no afecte al sistema completo. Integrados con prácticas de DevOps, optimizan recursos y reducen costos, posicionándose como una solución escalable y rentable [1]. El diseño de una arquitectura basada en microservicios no es trivial. Si las "fronteras" entre los microservicios no es la correcta se pueden generar "malos olores" que con el tiempo generan problemas en la evolución y el mantenimiento [2]. Al tomarse decisiones arquitecturales tempranamente (y con poca información) corremos el riesgo de hacerlo de manera equivocada dado que las decisiones de diseño arquitecturales son (como se afirma en [3]) aquellas que son más difíciles de modificar durante el ciclo de vida del software. Este problema se agrava cuando se utilizan enfoques ágiles de desarrollo [4] dado que los mismos rechazan la utilización de períodos largos de análisis y diseño (por ejemplo, arquitectural) previos al desarrollo. No es de extrañar, entonces, que muchas compañías hayan decidido dar marcha atrás en sus migraciones a microservicios y volver a centrar sus sistemas críticos en arquitecturas monolíticas [5]. Por estas razones (entre otras), se discute si en un sistema basado en MS es preferible diseñar los MS tempranamente o debe comenzarse como en un monolito para dividirse (luego, si fuera necesario) en MS [6]. En [7], como un argumento razonable, se indica que no se debería comenzar con MS sino con un monolito modular y, solo cuando (y sí) se convierte en un problema, descomponerlo en MS (usando técnicas existentes y documentadas). Este problema incluso afecta a desarrolladores más experimentados [8].

1.2. Metodologías Ágiles en el Contexto de los Microservicios

Existe otro problema, no menor, que tiene que ver con la evolución de los requerimientos y su impacto en el diseño de una arquitectura de MS. Si bien hay numerosos estudios que abordan el tema desde la perspectiva ágil, no identificamos en la literatura discusiones sólidas sobre las implicancias de esta evolución y las consecuencias de tomar las decisiones de diseño en las etapas incorrectas. Nos referimos a entender cuando se viola el principio de "no lo vas a necesitar" [9] o sufrir la falacia del gran diseño por adelantado [10]. En este contexto, identificamos las arquitecturas ágiles como una conciliación entre la agilidad y la arquitectura [11] permitiendo reducir el riesgo que genera dedicarle mucho tiempo a planificar y diseñar la arquitectura por adelantado. Varios autores han propuesto el uso de Domain-Driven Design (DDD) [12] como una estrategia adecuada para el descubrimiento de MS. Su uso podría reducir los malos olores mencionados anteriormente [13]. DDD propone basar el proceso de construcción de software en una comprensión profunda del modelo de negocios (en colaboración con los expertos del dominio) y de esa manera "descubrir" las abstracciones adecuadas, algunas de las cuales nos pueden ayudar a encontrar las fronteras entre MS. DDD aplicado en el contexto de las arquitecturas ágiles [11] permitiría reducir el riesgo de tomar decisiones inadecuadas tempranamente. Además, la combinación de DDD con prácticas ágiles consolidadas como Test-Driven Development [14] permitirían asegurar que el diseño emergente [15] se mantenga correcto. No obstante, se ha reportado [16] que no existe un soporte pragmático (en términos, por ejemplo, de heurísticas, reglas, etc) para guiar el reconocimiento ágil de microservicios en sistemas complejos. Los equipos de desarrollo enfrentan problemas para mantener consistencia entre modelos de dominio y su implementación. Además, la dificultad que existe para que desarrolladores y técnicos adquieran el conocimiento específico del dominio, valorando a las metodologías ágiles que alientan la participación de stakeholders en el proceso de desarrollo desde las etapas más tempranas. Adicionalmente, la naturaleza técnica de TDD, basada en un nivel de abstracción mucho más bajo (código) que el atacado por DDD, hace que exista un gap entre los conceptos apropiados para razonar sobre MS y los necesarios para aplicar esta técnica.

1.3. Inteligencia Artificial en el Descubrimiento y Diseño de Microservicios

Una de las mayores dificultades en el diseño es garantizar que los microservicios cumplan con principios fundamentales, como la cohesión interna y el bajo acoplamiento. Estos principios son esenciales para maximizar la eficiencia operativa y la flexibilidad del sistema, pero su aplicación práctica depende en gran medida de la experiencia del diseñador y de la calidad de los requisitos disponibles. En entornos donde los requisitos se presentan en forma textual y no estructurada, este desafío se amplifica, lo que aumenta el riesgo de errores de diseño y retrabajo[17]. La integración de IA, particularmente el procesamiento de lenguaje natural (PLN) y el aprendizaje profundo, representa una oportunidad única para abordar estas limitaciones. Estas tecnologías han demostrado su capacidad para extraer información estructurada de textos no estructurados, lo que las hace ideales para analizar documentos de requisitos [18]. Además, estudios recientes destacan la efectividad del clustering y otras técnicas de aprendizaje automático en la identificación de dominios funcionales y módulos coherentes [19].

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

2.1- Descubriento de Microservicios utilizando DDD y BDD

En este sub-proyecto buscamos investigar la utilización sistemática de Behavior-Driven Development (BDD) [20] como un complemento a DDD para descubrir, testear y soportar la evolución de MS. Técnicas como Test-Driven Development (TDD) [21] simplifican la refactorización del código en procesos ágiles, aunque suelen excluir a interesados como analistas y expertos del dominio. Behavior-Driven Development (BDD),

en cambio, amplía el alcance al incluir a todos los involucrados en el proceso. Basado en prácticas ágiles como TDD [26] y Domain-Driven Design (DDD), BDD utiliza un lenguaje ubicuo propuesto por Evans [12] y adaptado por Dan North [20] para describir requisitos de forma clara y accesible. Este enfoque permite transformar escenarios definidos por analistas en pruebas de aceptación automatizadas, facilitando la comprensión entre técnicos y expertos del dominio [22]. En entornos con requisitos cambiantes, BDD se presenta como una herramienta clave para construir características progresivamente y garantizar que el software cumpla con las expectativas del negocio [23]. El desarrollo de aplicaciones basadas en microservicios (MS) BDD es un área poco explorada, a pesar de que BDD está ganando popularidad por su enfoque ágil y centrado en el dominio, facilitando el uso de Domain-Driven Design (DDD). Sin embargo, no existen mecanismos adecuados para identificar MS de manera temprana utilizando los principios ágiles de BDD a partir de escenarios descritos en lenguaje específico de dominio [24].

2.2-Diseño de Microservicios usando IA generativa

Este sub-proyecto busca desarrollar un framework que utiliza técnicas de IA para automatizar el descubrimiento y diseño de microservicios desde requisitos textuales. Este enfoque no solo optimizaría el tiempo y esfuerzo necesario para el diseño, sino que también mejoraría la calidad arquitectónica al incorporar métricas objetivas (cohesión y bajo acoplamiento, alineación con los requisitos de negocios) y herramientas de validación automática del tipo de las existentes en TDD. Buscamos además validar el framework con casos reales.

3. RESULTADOS OBTENIDOS Y ESPERADOS

3.1 Resultados obtenidos

En el marco del sub-proyecto “Descubriento de Microservicios utilizando DDD y BDD”, se realizó un mapeo sistemático [25] sobre el descubrimiento de microservicios utilizando metodologías ágiles. Este trabajo revela una diversidad de enfoques y técnicas para identificar y definir los límites de los microservicios. La aplicabilidad práctica de estas técnicas en entornos empresariales reales sugiere que la implementación exitosa de microservicios requiere una adaptación continua y un enfoque colaborativo para abordar los cambios frecuentes en los requisitos. Asimismo, reveló que no existe gran cantidad de estudios versen sobre la creación de arquitecturas basadas en MS desde su concepción. Los estudiantes siguen trabajando con este enfoque, en búsqueda de definir procesos que permitan simplificar la identificación de MS. Por otro lado, en el contexto del doctorado en Informática de la UAI, se realizó un experimento para comparar enfoques en la identificación y diseño de microservicios (MS). Participaron 26 estudiantes divididos en 10 equipos, organizados en dos grupos: el Grupo A, que resolvió el problema sin seguir una metodología específica, y el Grupo B, que utilizó metodologías ágiles y Domain-Driven Design (DDD). El trabajo práctico consistió en identificar los MS necesarios para un escenario de gestión de manufactura, seguido de una segunda etapa en la que se introdujeron nuevos requisitos para analizar el impacto del cambio según el enfoque aplicado inicialmente. Al finalizar el experimento, se encuestó a los alumnos participantes, obteniendo los siguientes resultados: el 97% calificó el enunciado del problema como complejo o muy complejo. Respecto al nuevo requisito, el 57% tuvo que refactorizar su trabajo inicial, y el 87% lo consideró complejo o muy complejo. Entre quienes no usaron metodología, solo el 36% refactorizó, mientras que el 50% de los que usaron DDD y metodologías ágiles rediseñaron su trabajo. El análisis destaca que combinar un diseño inicial ligero con metodologías ágiles es clave para sistemas basados en microservicios (MS). Aunque el diseño inicial no elimina la necesidad de ajustes, facilita la identificación de límites contextuales y dominios clave, optimizando la granularidad y reduciendo riesgos. Los datos sugieren que el enfoque ágil y DDD requiere más refactorización, pero ofrece un marco sólido que equilibra planificación y adaptabilidad, reduciendo la deuda técnica y

mejorando escalabilidad y alineación entre objetivos técnicos y de negocio.

En el subproyecto 2 "Diseño de Microservicios usando IA generativa", se realizó un Mapeo Sistemático de la Literatura (MSL) para identificar enfoques existentes en el diseño de microservicios mediante IA y metodologías ágiles [10]. Se analizaron distintas estrategias de diseño, comparando el uso de IA generativa frente a métodos tradicionales. Los resultados indican que la IA generativa puede reducir significativamente el tiempo de diseño y mejorar la identificación de límites de servicio.

3.2. Resultados esperados

Actualmente, estamos trabajando en otro mapeo sistemático que pretende estudiar la relación que existe entre DDD y BDD y la manera que estas técnicas pueden coexistir para favorecer la identificación y posterior mantenibilidad de microservicios en el contexto de arquitecturas ágiles. En el sub-proyecto 2 estamos comenzando a utilizar herramientas de procesamiento de lenguaje natural para procesar documentos de requisitos y extraer sub-dominios, entidades claves y acciones. Buscaremos agrupar las entidades extraídos en módulos cohesivos (en los sub-dominios) usando algoritmos de clustering junto a redes neuronales profundas.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de investigación está compuesto por 5 estudiantes de grado y posgrado de la UAI, quienes están desarrollando sus respectivas PPS, trabajos finales y tesis en temas relacionados con el proyecto. Dos estudiantes de grado (Ingeniería en Sistemas) realizaron su PPS en el marco del proyecto y continúan investigando sobre la misma temática. Por otro lado, dos doctorandos estan realizando sus tesis. Por un lado. estudiando la identificación ágil de microservicios utilizando DDD y BDD y por el otro explorando el uso de inteligencia artificial en el descubrimiento y diseño de microservicios. Por último, un estudiante de maestría está realizando su tesis en el marco del uso de DDD y BDD como herramientas conjuntas para la comprensión de requisitos

5- BIBLIOGRAFÍA

[1] L. De Lauretis, "From monolithic architecture to microservices architecture," in *2019 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW)*, 2019, pp. 93–96.

[2] D. Taibi and V. Lenarduzzi, "On the definition of microservice bad smells," *IEEE Softw*, vol. 35, no. 3, pp. 56–62, 2018.

[3] M. Fowler, *Patterns of enterprise application architecture*. Addison-Wesley, 2012.

[4] P. Kruchten, "Software architecture and agile software development: a clash of two cultures?," in *Proceedings of the 32nd ACM/IEEE International Conference on Software Engineering-Volume 2*, 2010, pp. 497–498.

[5] R. Su, X. Li, and D. Taibi, "From Microservice to Monolith: A Multivocal Literature Review," *Electronics (Basel)*, vol. 13, no. 8, p. 1452, 2024.

[6] J. P. S. Cardoso, "A guide for microservices in greenfield projects," Instituto Politecnico do Porto (Portugal), 2021.

[7] J. Lewis and M. Fowler, "Microservices - a definition of this new architectural term." Accessed: Sep. 03, 2024. [Online]. Available: https://martinfowler.com/articles/microservices.html

[8] M. Fowler, "Monolith First." Accessed: Jan. 05, 2025. [Online]. Available: https://martinfowler.com/bliki/MonolithFirst.html

[9] M. Fowler, "YAGNI." Accessed: Jan. 05, 2025. [Online]. Available:

https://martinfowler.com/bliki/Yagni.html

[10] O. Zimmermann, "Microservices tenets: Agile approach to service development and deployment," *Computer Science-Research and Development*, vol. 32, pp. 301–310, 2017.

[11] M. G. Waterman, "Reconciling agility and architecture: a theory of agile architecture," Open Access Te Herenga Waka-Victoria University of Wellington, 2014.

[12] E. Evans, *Domain-driven design: tackling complexity in the heart of software*. Addison-Wesley Professional, 2004.

[13] H. Vural and M. Koyuncu, "Does domain-driven design lead to finding the optimal modularity of a microservice?," *IEEE Access*, vol. 9, pp. 32721–32733, 2021.

[14] K. Beck, *Test driven development: By example*. Addison-Wesley Professional, 2022.

[15] S. Blair, R. Watt, and T. Cull, "Responsibility-driven architecture," *IEEE Softw*, vol. 27, no. 2, pp. 26–32, 2010.

[16] C. Zhong *et al.*, "Domain-driven design for microservices: An evidence-based investigation," *IEEE Transactions on Software Engineering*, 2024.

[17] D. Taibi, V. Lenarduzzi, and C. Pahl, "Processes, motivations, and issues for migrating to microservices architectures: An empirical investigation," *IEEE Cloud Computing*, vol. 4, no. 5, pp. 22–32, 2017.

[18] D. Narváez, G. Rossi, and N. Battaglia, "Aplicación de inteligencia artificial en el diseño de microservicios: Un mapeo sistemático de la literatura," in *Libro de Actas-30° Congreso Argentino de Ciencias de la Computación-CACIC 2024*, Facultad de Informática (UNLP), 2024, pp. 639–648. [Online]. Available: http://dx.doi.org/10.35537/10915/172755

[19] M. A. I. T. SAID, A. EZZATI, S. MIHI, and L. BELOUADDANE, "Microservices adoption: An industrial inquiry into factors influencing decisions and implementation strategies," *International Journal of Computing and Digital Systems*, vol. 15, no. 1, pp. 1417–1432, 2024.

[20] D. North and others, "Introducing BDD," *Better Software*, vol. 12, p. 7, 2006.

[21] A. Hellesoy, S. Tooke, and M. Wynne, *The cucumber book: behaviour-driven development for testers and developers*. The Pragmatic Bookshelf, 2017.

[22] T. Tuglular, D. E. Coşkun, Ö. Gülen, A. Okluoğlu, and K. Algan, "Behavior-Driven Development of Microservice Applications," *International Journal of Computers*, vol. 15, pp. 130–137, 2021, doi: 10.46300/9108.2021.15.20.

[23] J. F. Smart and J. Molak, *BDD in Action: Behavior-driven development for the whole software lifecycle*. Simon and Schuster, 2023.

[24] S.-P. Ma, Y. Chuang, C.-W. Lan, H.-M. Chen, C.-Y. Huang, and C.-Y. Li, "Scenario-based microservice retrieval using Word2Vec," in *2018 IEEE 15th International Conference on e-Business Engineering (ICEBE)*, 2018, pp. 239–244.

[25] N. Battaglia, A. N. Garcia, and A. Congiusti, "Descubrimiento de Microservicios en Metodologías Ágiles: un mapeo sistemático de la literatura," in *XXX Congreso Argentino de Ciencias de la Computacion*, 2024, pp. 1506–1510.

Ciudades Inteligentes Sostenibles. Gobernanza Digital, IA en Ingeniería de Software, Buenas prácticas y Calidad de la IA.

Ariel Pasini, Marcos Boracchia,
Rocío Muñoz, Kristian Petkoff Bankoff , Ailen Panigo, Juan Ignacio Torres, Patricia Pesado

Instituto de Investigación en Informática LIDI (III-LIDI)
Facultad de Informática – Universidad Nacional de La Plata
50 y 120 La Plata Buenos Aires
Centro Asociado CIC
526 e/ 10 y 11 La Plata Buenos Aires
(apasini, marcosb, rmunoz, kpb, apanigo, jitorres, ppesado) @lidi.info.unlp.edu.ar

RESUMEN

El III-LIDI (Instituto de Investigación en Informática LIDI) posee un subproyecto centrado en cuatro ejes principales
1. Ciudades inteligentes sostenibles; 2. Gobernanza digital y gobierno abierto; 3. Buenas prácticas y calidad; y 4. Inteligencia Artificial en Ingeniería de Software. Para cada eje se proponen un conjunto de objetivos para analizar durante este año y se describen los resultados obtenidos hasta el momento.

Palabras Claves
Ingeniería de Software – Gobernanza Digital – Ciudades Inteligentes – Transformación Digital – Inteligencia Artificial

CONTEXTO

Esta línea de investigación se enmarca en el proyecto "Diseño, desarrollo y evaluación de sistemas en escenarios híbridos para áreas clave de la sociedad actual: educación, ciudades inteligentes y gobernanza digital" (2023-2026), UNLP, Ministerio de Educación de la Nación, en particular el subproyecto "Ciudades inteligentes sostenibles. Gobernanza digital. Buenas prácticas y calidad". Y posee una estrecha interacción con los otros proyectos desarrollados en el Instituto que se relacionan con Inteligencia de Datos y Computación de Alto Desempeño y Distribuida.
Asimismo, la temática se encuentra abordada en proyectos aprobados por la Facultad de Informática UNLP (2024-2025).
El Instituto posee diversos acuerdos de cooperación con varias Universidades de Argentina y del exterior y con empresas privadas del sector, interesadas en mejorar sus procesos de desarrollo.
Desde esta línea, el III-LIDI participa en proyectos internacionales que abordan los temas de Transformación Digital (TD), Ciudades Digitales e Inteligencia Artificial (IA):
-Actualmente se trabaja en el proyecto *"UNI-UEAR - Alianza Universitaria Argentina Europea para la Transformación Digital"* cofinanciado por la Unión Europea. El Proyecto UNI UEAR es ejecutado por el Consejo Interuniversitario Nacional (CIN) en colaboración con un consorcio de cinco universidades nacionales UBA, UNLP, UNQ, UNS, UNSaM y cuatro instituciones europeas: European University Institute (Italia), Gdansk University of Technology (Polonia), Technical University of Delft (Países Bajos), University for Continuous Education Krems (Austria). Tiene como

objetivo principal contribuir a la promoción de los derechos digitales en Argentina; y en particular, busca fortalecer la creación y difusión de conocimiento aplicado para la transformación digital; y desarrollar alianzas entre la academia, los sectores público y privado y la sociedad civil para la transformación digital.
-Además, el Instituto posee una importante presencia en el "Consorcio de I+D+I en Cloud Computing-Big Data & Emergent Topics" (CC-BD&ET) con la participación de grupos de investigación de Argentina, España, Chile, entre otros.
- Se encuentra en evaluación en el marco de una convocatoria del programa ERASMUS+ el proyecto internacional "REACT4CITIES Resilience, Empowerment, and Al-powered Capacity-building for Transformation in Cities and Communities" en el que participan las universidades de UWK (Austria), GUT (Polonia), TUDelft, (Países Bajos), PUCRS, UFRJ y FGV-SP(Brasil), UNLP, UNS, y UNQ (Argentina).
- Se encuentra en evaluación el proyecto internacional "INSPIRE - Innovative Networks for Smart Policies, Integration, Research, and Engagement in Smart Cities and Governance" de intercambio de recursos humanos coordinado por UWK (Austria) donde el Instituto participa.

1. INTRODUCCION

La línea de investigación se focaliza en 4 ejes. El primero basado en el concepto de ciudades inteligentes sostenibles (CIS), la generación de herramientas para capacitar a la comunidad en el desarrollo de CIS y la transformación digital. El segundo eje está directamente relacionado con el concepto de CIS. Se profundiza en el uso de la tecnología en la gobernanza digital, gobierno y datos abiertos. El tercer eje se enfoca en distintos lineamientos referidos a la generación de buenas prácticas, la mejora de procesos relacionados a los organismos públicos, y la calidad en los emergentes sistemas de IA. Por último el cuarto eje, busca investigar la injerencia de la utilización de herramientas basadas en modelos de lenguajes de gran tamaño (LLM) de inteligencia artificial como por ejemplo, ChatGPT, Llama, etc. en los procesos definidos en la Ingeniería de Software

Eje 1 - Ciudades inteligentes sostenibles

El proyecto CAP4CITY define una CIS como: "...una ciudad innovadora que utiliza Tecnologías de Información y Comunicación (TIC) y otros medios para mejorar la calidad de vida, la eficiencia de las operaciones y los servicios urbanos; y la competitividad, al tiempo que se garantiza que satisfaga las necesidades de las generaciones presentes y futuras con respecto a los aspectos económicos, sociales, ambientales y culturales."[1] - [5].
En este contexto las CISs se componen de cinco pilares: "Social", "Económico", "Ambiental", "Gobernanza" e "Infraestructura urbana".
Las ciudades proveen servicios públicos mediante el uso TICs a su comunidad, si estos servicios son orientados a la innovación y la sostenibilidad, se lo considera un servicio CIS.
Asistiendo al ciudadano en la incorporación de una forma más sencilla al proceso de transformación digital

Dentro de los servicios "Ambientales", podemos mencionar como ejemplos: el uso de gemelos digitales en la producción agrícola que permiten gestionar la toma de decisiones en el uso de equipos de riego haciendo un uso racional del agua, y el modelo predictivo de residuos electrónicos desde la perspectiva de le ingeniería ambiental.

Eje 2 - Gobernanza Digital y Gobierno Abierto

Continuando con los pilares de las CIS mencionados anteriormente, dentro del pilar de Gobernanza podemos destacar el concepto de gobernanza digital, este

concepto, se basa en el uso de las TICs para mejorar la forma de gobierno, la entrega de servicios públicos y facilitar las interacciones entre el gobierno y el ciudadano [3], [5], [7], [8].
En este contexto podemos mencionar el concepto de gobernanza universitaria, en el cuan es importante medir el nivel de satisfacción del ciudadano universitario, en particular en el uso de herramientas digitales como SUDOKU, SIU, PILAGA, etc….

Eje 3- Buenas Prácticas y Calidad de I.A.

Una buena práctica se define como un conjunto de acciones que han sido efectivas en un determinado contexto, y se espera obtener el mismo resultado en contextos similares. Estas buenas prácticas se pueden encontrar tanto en tareas que se realizan de manera cotidiana como en la prestación de servicios. Tomando como premisa estas buenas prácticas, se puede mejorar la prestación de servicios de una organización. Desde diferentes puntos de vista, como por ejemplo el producto, el proceso de desarrollo o el servicio prestado, entre otros. Estos puntos de vista suelen tener asociados estándares internacionales que facilitan la identificación de buenas prácticas.
El avance de la TD en particular el crecimiento en el uso de herramientas de IA requiere de analizar desde diferentes perspectivas el nivel de calidad con el que se desarrollan y prestan los servicios que usan ese tipo de tecnología.

Eje 4 - IA en Ingeniería de Software

La Inteligencia Artificial (IA) se ha fusionado progresivamente con las prácticas de ingeniería y desarrollo de software, marcando avances significativos en el campo. Esta integración tiene como objetivo mejorar la eficiencia, la creatividad y la automatización de los procesos de ingeniería de software, desde el diseño hasta el mantenimiento. La relación entre la IA y el desarrollo de software abarca varias facetas, incluida la automatización de tareas rutinarias, el análisis de datos, el apoyo a las decisiones y el desarrollo de sistemas inteligentes. La viabilidad de que la IA ayude en el desarrollo de software está bien justificada por las direcciones actuales y los usos de casos destacados en numerosos estudios.
La IA contribuye al desarrollo de software al automatizar largas tareas rutinarias, como la depuración y la documentación. Barenkamp, Rebstadt y Thomas [17] (2020) enfatizan el papel de la IA para acelerar los procesos de desarrollo, lograr reducciones de costos y ganancias de eficiencia. Las herramientas de IA multiplican eficazmente el potencial creativo de los desarrolladores humanos al hacerse cargo de tareas repetitivas, lo que les permite centrarse en actividades de resolución de problemas más complejas.
El análisis estructurado de grandes conjuntos de datos para descubrir patrones y grupos de información novedosos es otra área donde la IA puede asistir en el desarrollo de software. Salehi & Burgueño [18]mencionan que se han adoptado cada vez más técnicas como el aprendizaje automático, el reconocimiento de patrones y el aprendizaje profundo, lo que demuestra la capacidad de la IA para manejar incertidumbres y resolver problemas complejos que van más allá de los métodos tradicionales. Estas soluciones basadas en IA brindan alternativas para los parámetros de diseño de ingeniería cuando no es posible realizar pruebas directas, lo que genera importantes ahorros de tiempo y esfuerzo.
La integración de la IA en el desarrollo de software enfrenta desafíos como la necesidad de que los ingenieros de software adquieran nuevas habilidades y se adapten a los constantes cambios de roles. Los ingenieros de software deben ser expertos no solo en programación sino también en la utilización de herramientas de inteligencia artificial para el desarrollo. La evolución requiere un aprendizaje y una

adaptación continuos para aprovechar plenamente la IA en las prácticas de ingeniería de software [19].
La etapa de de elicitacion y análisis de requerimientos son cruciales en el proceso de desarrollo y los modelos de IA se presentan como una herramienta para facilitar la generación de entrevistas y cuestionarios para el entrevistador y luego en la consulta de historias de usuario por los analistas.

2. LÍNEAS DE INVESTIGACIÓN y DESARROLLO

- Análisis de una CIS desde diferentes aspectos (por ejemplo, gemelos digitales en la producción agrícola, modelos de predicción de residuos electrónicos, modelos de madurez de satisfacción de ciudadanos universitarios, formación en Innovación y co-creación de servicios públicos)
- Análisis del impacto de la TD en los distintos sectores de la sociedad.
- Desarrollo de prototipos para Voto electrónico presencial y Voto por Internet. Arquitecturas adaptadas a la legislación vigente en distintos organismos.
- Análisis y estudio de normas y modelos orientados a la Calidad relacionados con IA
- Análisis, discusión y estudio de técnicas y herramientas de mejoras de proceso en el desarrollo de software.
- Análisis y discusión del uso de IA en el proceso de desarrollo del software.

3. RESULTADOS OBTENIDOS/ ESPERADOS

- Esta en marcha la primera cohorte de la Maestría en Gestión y Tecnología de Ciudades Inteligentes (UNLP-UNS) que fue definida en el contexto del Proyecto CAP4CITY.
- Se avanza en el desarrollo de un modelo de evaluación de satisfacción del ciudadano universitario, con una tesis de doctorado en desarrollo en conjunto con la Facultad de Humanidades de la UNLP.
- Se avanza en el desarrollo de un modelo predictivo de residuos electrónicos desde la perspectiva de la ingeniería ambiental, con una tesis doctoral en curso.
- Se avanza en un modelo basado en gemelos digitales aplicados al sector agrícola, con el inicio de una tesis de doctorado.
- Se realizaron acciones de consultoría y asesoramiento en organismos públicos y privados, en particular se aplicó el Voto por Internet en distintos organismos.
- Se realizaron la elecciones del claustro de estudiantes de la Facultad de informática del 2024 utilizando los equipos de voto electrónico desarrollados recientemente.
- Se mantiene el Sistema de Gestión de la Calidad de la Facultad (SIGCFI).
- En el marco del proyecto UNIUEAR, el Instituto está desarrollando tres cursos:
    1. Tecnologías para la TD
    2. Gobernanza de datos
    3. Gestión de la Innovación

y se trabaja en la definición de proyectos de investigación en temas relacionados, en conjunto con otras universidades del país.
- Se participó en la definición del proyecto internacional presentado a la convocatoria del programa ERASMUS+ “REACT4CITIES Resilience, Empowerment, and Al-powered Capacity-building for Transformation in Cities and Communities".
- Se analizaron las posibilidades de aplicar IA en el proceso de desarrollo de software.

4. FORMACION DE RECURSOS HUMANOS

- Se dirigen tesis de postgrado, tesinas de grado y PPS en temas del área.
- Se dirige la Maestría en Gestión y Tecnología de Ciudades Inteligentes (UNLP-UNS).
- Se participa en el dictado de asignaturas/cursos de grado/postgrado en la Facultad de Informática de la UNLP y en otras universidades del país.

5. BIBLIOGRAFIA

[1] E. Estevez, “Innovació n en Servicios P ú blicos Digitales.”
[2] ITU-T, “An overview of smart sustainable cities and the role of information and communication technologies,” 2014.
[3] R. Muñoz, A. Pasini, and P. Pesado, “Innovación en el Sector Público para Ciudades Inteligentes Sostenibles.”
[4] C. Sánchez, *Innovación pública:* .
[5] S. A. Chun, S. Shulman, R. Sandoval, and E. Hovy, “Government 2.0: Making connections between citizens, data and government,” *Inf. Polity*, vol. 15, no. 1–2, pp. 1–9, 2010.
[6] A. E. Millat, “Ciudades Resilientes Formando un Mundo Digital,” 2020.
[7] S. Esponda, A. Pasini, M. Boracchia, J. Calabrese, R. Muñoz, J. Santiago Preisegger, and P. Pesado, “Fortalecimiento de la calidad en procesos de software y Gobernanza Digital,” *Sedici.Unlp.Edu.Ar*, pp. 574–578.
[8] J. S. Preisegger, A. Greco, A. Pasini, M. Boracchia, and P. Pesado, “Marco de vinculación de datos abiertos aplicado al contexto de datos medioambientales,” pp. 684–694.
[9] A. Naser and A. Ramirez, “Plan de gobierno abierto. Una hoja de ruta para los Gobiernos de la Región,” *CEPAL - Manuales*, vol. 81, p. 80, 2017.
[10] A. Pasini, J. S. Preisegger, and P. Pesado, *Open Government Assessment Models Applied to Province’s Capital Cities in Argentina and Municipalities in the Province of Buenos Aires*, vol. 995. 2019.
[11] C. Calderón and S. Lorenzo, *Open Government. Gobierno Abierto*. 2010.
[12] N. Noy, M. Burgess, and D. Brickley, “Google dataset search: Building a search engine for datasets in an open web ecosystem,” in *The Web Conference 2019 - Proceedings of the World Wide Web Conference, WWW 2019*, 2019, pp. 1365–1375.
[13] J. Calabrese, R. Muñoz, A. Pasini, S. Esponda, and M. Boracchia, “Asistente para la evaluación de características de calidad de producto de software propuestas por ISO / IEC 25010 basado en métricas definidas usando el enfoque GQM,” pp. 660–671, 2017.
[15] A. Pasini, E. Estévez, P. Pesado, and M. Boracchia, “Una metodología para evaluar la madurez de servicios universitarios,” *Proc. Congr. XXII Congr. Argentino Ciencias la Comput. Congr. Argentino Ciencias la Comput.*, pp. 636–646, 2016.
[16] A. Pasini, J. S. Preisegger, and P. Pesado, “Modelos de evaluación de gobiernos abiertos , aplicado a los municipios de la provincia de Buenos Aires,” *XXIV Congr. Argentino Ciencias la Comput.*, vol. XXIV, pp. 0–10, 2018.
[17] Barenkamp, M., Rebstadt, J., & Thomas, O. (2020). Applications of AI in classical software engineering. AI Perspectives, 2, 1-15. https://doi.org/10.1186/s42467-020-00005-4.
[18] Salehi, H., & Burgueño, R. (2018). Emerging artificial intelligence methods in structural engineering. Engineering Structures. https://doi.org/10.1016/J.ENGSTRUCT.2018.05.084.
[19] Necula, S. (2023). Artificial Intelligence Impact On The Labour Force - Searching For The Analytical Skills Of The Future Software Engineers. ArXiv, abs/2302.13229. https://doi.org/10.48550/arXiv.2302.13229

CREACIÓN DE UN MODELO INTEGRAL DE GESTIÓN DEL CONOCIMIENTO PARA LAS PYMES ARGENTINAS

Straccia, Luciano; Pollo-Cattaneo, María F.; Baruzzini, Brenda; Berro, Camila

Universidad Tecnológica Nacional, Facultad Regional Buenos Aires, Grupo GEMIS.BA, Ciudad Autónoma de Buenos Aires, Argentina
lstraccia@frba.utn.edu.ar, flo.pollo@gmail.com, bbaruzzini@frba.utn.edu.ar, camiberro.cb@gmail.com

RESUMEN

La gestión del conocimiento (GC) aborda la creación y el uso del conocimiento en las organizaciones desde un enfoque multidisciplinar. El Grupo GEMIS.BA, a través de diversos proyectos, se propone analizar la GC desde sus diversos enfoques y presentar un modelo integral para su aplicación concreta en las organizaciones. Este trabajo presenta los diferentes enfoques, los proyectos de investigación desarrollados en dicho marco, los resultados obtenidos y un nuevo proyecto de investigación y desarrollo. Los proyectos (finalizados o en desarrollo) del grupo de investigación han tratado sobre la indagación del grado de implementación de GC en las pequeñas y medianas fábricas de software; análisis de modelos de representación del contenido y propuestas de tecnologías y arquitecturas tecnológicas para la implementación de GC en las organizaciones; definición de un modelo de medición de GC y un observatorio tecnológico que favorezca la visualización del estado de aplicación de la GC en la Ciudad de Buenos Aires. Este nuevo proyecto aporta una visión integral de los diferentes aspectos de la GC.

Palabras clave: conocimiento, gestión del conocimiento, metodología, tecnología, medición.

CONTEXTO

La Universidad Tecnológica Nacional (UTN) posee un Programa de Investigación, Desarrollo e Innovación, de Sistemas de Información e Informática. En la Facultad Regional Buenos Aires (UTN.BA), en el marco del Departamento de Ingeniería en Sistemas de Información, se llevan adelante proyectos de investigación y desarrollo. Allí desarrolla sus actividades, entre otros grupos de investigación, el grupo GEMIS.BA que trabaja en el campo de la Ingeniería en Sistemas de Información, Ingeniería de Software e Ingeniería del Conocimiento.

El Grupo GEMIS.BA realiza, en el marco de las temáticas objeto de este trabajo, diversos proyectos asociados a la línea de Gestión del Conocimiento con proyectos de investigación y desarrollo y actividades de docencia e investigación en diferentes instituciones contando con trabajos finales de especialización, tesis de maestría y un plan de tesis doctoral en proceso.

1. INTRODUCCIÓN

La Gestión del Conocimiento aborda la creación y uso del conocimiento en las organizaciones [1]. El conocimiento forma parte de la jerarquía DIKW (datos, información, conocimiento y sabiduría) [2] y es "la mezcla de creencias cognitivas (..), perspectivas, juicios, metodologías, (…) experiencias y expectativas realizadas sobre un objeto, que son adaptadas y potenciadas por la mente de un individuo (conocedor)" [3]. DIKW puede considerarse desde una perspectiva constructivista donde “conciben (sus conceptos) como una pirámide en

la que están jerarquizados de menor a mayor y son prerrequisitos el uno del otro en términos de conversión" [4]. En [4] presentan también el enfoque representativo que invierte la pirámide y "el conocimiento es el punto de partida que, al ser articulado, verbalizado y estructurado, llega a ser información, la que posteriormente se torna dato una vez que se atomiza y se le asigna una determinada representación e interpretación en una estructura semántica".

La gestión del conocimiento tiene fuerte presencia en trabajos académicos según se desprende de fuentes internacionales académicas (DBLP, ACM, DOAJ y World Wide Science) en búsquedas con "knowledge management" como palabra clave. Sin embargo, al incorporar "model" en las palabras claves esa presencia se reduce. La circulación de información respecto a gestión del conocimiento es alta, pero su especificidad en modelos parece resultar más baja. También debe tenerse en cuenta que el concepto gestión del conocimiento es utilizado en algunos campos con acepciones diferentes a las consideradas desde el campo de la gestión organizacional (destacándose fuertemente su uso en artículos del campo de la educación para referirse a aspectos vinculados a la enseñanza o al aprendizaje). Para Sinisterra-Nuñez y otros [5] "es importante señalar, aunque gestión del conocimiento es un término de uso generalizado [6] existen evidencias de cómo el significado y los enfoques que del mismo se tienen en el contexto de las organizaciones es muy diverso".

La incorporación de los conceptos de "modelo integrado" (o su variante "modelo integral") en las búsquedas muestra también una merma relevante en sus apariciones. Pueden entenderse estos resultados como una preocupación general sobre la gestión del conocimiento aunque ante la necesidad de una mayor especificidad la magnitud de los trabajos se reduce, encontrando dificultades para hallar publicaciones con una mirada integral de la gestión del conocimiento. La gestión del conocimiento es un área emergente que aún requiere de muchas investigaciones e innovación para adaptarse a los distintos rubros y explotar realmente los beneficios del conocimiento en la sociedad actual [7].

En [7] se analizan diferentes modelos de Gestión de Conocimiento; allí se puede observar que algunos modelos sólo definen una estructura global, mientras que la mayoría sólo propone actividades para la gestión del conocimiento. Una visión integral de la gestión del conocimiento (que englobe diferentes aspectos) no es hallada en ninguno de estos modelos. Esta visión más integral de la gestión del conocimiento viene siendo trabajada por el grupo de investigación en los últimos años teniendo en cuenta diferentes vistas: 1) individuos o personas, 2) aspectos organizativos, 3) actividades y procesos, 4) tecnología y representación y 5) medición [8]. A continuación, se presentan algunos antecedentes del grupo de investigación sobre las diferentes vistas.

En [7] se presentan y analizan diferentes modelos, en [9] se realiza una comparación entre modelos de GC buscando similitudes y aportaciones diferenciales e identificando fases requeridas para toda metodología de GC y en [10] se analizan las diferentes actividades para la GC organizándolas según cuatro perspectivas: individual, repositorio, de equipo e identificación, creación y adquisición.

En [11;12] se realiza un mapeo de literatura acerca de la tecnología para la GC obteniendo diferentes categorías. En [11] también se analizan arquitecturas tecnológicas. Inicialmente se analizaron los modelos de arquitectura de integración tecnológica de Kerschberg, Zavala-Zavala y de Jofré. En [13] se realiza un análisis de las arquitecturas de Ovum, Tiwana y Kerschberg y, considerando las capas que presentan, se propone una arquitectura de 6 capas: presentación, acceso y autenticación, gestión del conocimiento, taxonomías y mapas del conocimiento, gestión de la información y

fuentes y repositorios. En [14] se incrementa el análisis de arquitecturas a través de una amplia revisión bibliográfica.

Pérez López Portillo [15] afirma que, debido a su naturaleza compleja, la medición de la GC es uno de los temas menos desarrollados o investigados y que "es muy importante establecer medidas de rendimiento en las diferentes etapas de la implementación de la GC, e incluso desde el principio para poder identificar su eficacia". En [16] se presenta un modelo de madurez para medir la GC. Todos los modelos de madurez se organizan en 3 componentes: niveles de madurez, áreas de procesos clave y los descriptores para cada área de proceso clave en cada nivel

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

En el año 2018 se llevó a cabo el primer proyecto de investigación y desarrollo (PID) del Grupo GEMIS.BA asociado a GC en la Facultad Regional Buenos Aires de la Universidad Tecnológica Nacional. Este proyecto buscaba indagar cuál era el grado de implementación de GC en las pequeñas y medianas fábricas de software del Área Metropolitana de Buenos Aires (AMBA), cuáles eran los problemas presentes en esta implementación y qué modelos se podrían proponer para que esas organizaciones pudieran implementarlos (PID UTN, 2018-2020). Este proyecto obtuvo resultados construyendo un modelo general pero que se centraba especialmente en las actividades, desconociendo (o minimizando) otros aspectos.

A partir de los resultados obtenidos en el proyecto anterior y atendiendo a la necesidad de indagar en algunos aspectos particulares de la GC, en el año 2021 se iniciaron proyectos asociados a aspectos tecnológicos y a la medición de algunos de esos aspectos. En la UTN.BA se realizó un proyecto que busca analizar modelos de representación del conocimiento y proponer tecnologías y arquitecturas tecnológicas para la implementación de GC en las organizaciones (PID UTN 2021-2024). Además, entre los años 2021 y 2023 uno de los investigadores del Grupo dirigió un proyecto en la Universidad Argentina de la Empresa (UADE) cuyo objetivo fue definir un modelo de medición de GC construyendo indicadores sobre las tecnologías para la medición de GC y proponiendo un modelo de implementación de los indicadores en las organizaciones (PID UADE, 2021-2023). Finalmente, en el año 2023 se inició el proyecto de desarrollo de un observatorio tecnológico para relevar y monitorear el estado de aplicación de la gestión del conocimiento en organizaciones de la Ciudad de Buenos Aires (PID UTN, 2023-2025).

Los proyectos de investigación mencionados abordan diferentes vistas de la gestión del conocimiento brindando solidez y pericia para el proyecto que es presentado en este trabajo.

3. RESULTADOS OBTENIDOS/ESPERADOS

Las diferentes vistas mencionadas en la introducción han requerido un trabajo específico sobre cada una de ellas, con diferentes trabajos que fueron presentados en dicha sección. Esta segmentación en diferentes vistas aporta una mirada específica que ha permitido realizar aportes sustanciales a la GC. Sin embargo, también dicha segmentación abandona una visión integral que produzca sinergia entre las diferentes perspectivas. La construcción de dicha sinergia es la búsqueda de este proyecto.

En particular en este proyecto el grupo de investigación posee un especial interés en enfocarse en las pequeñas y medianas empresas de la Argentina, considerando sus características y sus dificultades. La pequeña y mediana

empresa se ha constituido como un objeto de estudio relevante en los últimos años [17]. Es notorio que la realidad de las empresas más pequeñas es diferente a aquellas de mayor tamaño encontrándose con algunas dificultades.

Algunos de los problemas que se presentan en este tipo de empresas es el excesivo nivel de centralización de la gestión en el empresario, niveles de inversión incremental que no permiten alcanzar un adecuado nivel de actualización tecnológica y baja demanda de servicios de apoyo [18]. También se encuentran deficiencias en la gestión, problemas de falta de asociatividad, influencias negativas de las variables del entorno y la falta de crédito y apoyo gubernamental. Las dificultades de acceso a la tecnología son presentadas en diversos trabajos [17], incluso muchas organizaciones no cuentan con equipos de gestión tecnológica. En un trabajo reciente desarrollado en España [19] se puede encontrar el acceso a la tecnología como un factor problemático para casi un tercio de las organizaciones pequeñas y medianas y la pretensión de un mayor fomento a la transformación digital, facilitación de procesos de cooperación y mayor vinculación con otras organizaciones, todos aspectos vinculables a la gestión del conocimiento. De similar forma, se presentan problemas para Colombia en donde los estándares para adquisición y desarrollo están dirigidos a grandes empresas con realidades diferentes a las organizaciones pequeñas del país y que se requiere un modelo de software para esas organizaciones [20]. En Argentina, las pequeñas y medianas empresas representan el 99,4% del total de empresas del país y emplean al 64% de los asalariados registrados [21]. El enfoque sobre las Pymes argentinas permite un amplio escenario posible de transferencia.

Por lo presentado anteriormente, las siguientes preguntas guían el proyecto de investigación: ¿cómo se articulan las diferentes vistas de la gestión del conocimiento para constituir un modelo integral? y ¿qué características debe considerar un modelo de gestión del conocimiento para su implementación en pequeñas y medianas empresas de la Argentina?

El objetivo general del proyecto es “definir un modelo integral de gestión del conocimiento aplicable a las pequeñas y medianas empresas de la Argentina” y sus objetivos específicos son: 1) identificar las características y limitaciones de las pequeñas y medianas empresas argentinas; 2) diseñar un modelo integral de gestión del conocimiento; 3) proponer una metodología para la implementación del modelo integral de gestión del conocimiento; 4) diseñar manuales para la implementación del modelo integral de gestión del conocimiento.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo propuesto se encuentra conformado por investigadores formados, investigadores de apoyo y alumnos de la carrera de Ingeniería en Sistemas de Información. Esta línea de trabajo busca tanto la obtención de nuevos conocimientos como la motivación de los implicados para que asciendan dentro de la carrera de investigadores, además de fomentar la aplicación de las temáticas en su propia actividad profesional.

Además el equipo cuenta con investigadores estudiantes que realizan sus primeras participaciones en el ámbito de la investigación y cuya experiencia en el proyecto propuesto contribuirá especialmente al inicio de trayectos de formación académica. Se espera la incorporación de nuevos becarios alumnos y graduados a partir de las diferentes propuestas de becas de investigación que son llevadas a cabo por la Universidad y la Facultad.

Se espera el desarrollo de por lo menos 2 tesis de la Maestría en Ingeniería en Sistemas de Información de UTN-FRBA. Además se prevé

el inicio de por lo menos 2 trabajos finales de la Especialidad en Ingeniería en Sistemas de Información (UTN-FRBA). Finalmente, resultados de este proyecto se prevé que sean utilizados en el marco de una tesis doctoral de Doctorado en Ingeniería.

5. BIBLIOGRAFÍA

[1] Farfán Buitrago, D.; Garzón Castrillón, M. La gestión del conocimiento, Bogotá: Editorial Universidad del Rosario, 2006. 109 p.

[2] Ackoff, R. (1989). From data to wisdom. Journal of Applied Systems Analysis.

[3] Díaz, M.; Millán, J. (2013). Gestión del Conocimiento y Capital Intelectual a través de modelos universitarios. Económicas CUC.

[4] Arias Perez, J.; Aristizabal Boero, C. (2011). El dato, la información, el conocimiento y su productividad en empresas del sector público de Medellín. En Semestre Económico, Universidad de Medellín.

[5] Sinisterra-Nuñez, A.; Osorio-Bayter, L.; Gabalán-Coello, J. (2019). Ruta de aprendizaje de gestión del conocimiento en fundaciones empresariales. En Económicas CUC, vol. 40 no. 2, pp. 70-86, Julio - Diciembre, 2019.

[6] Piorun, D. (2018). Introducción a la gestión del conocimiento. Buenos Aires: Universidad de Buenos Aires

[7] Maulini, A (2021). Propuesta de un modelo de Gestión del Conocimiento para fábricas de software. Tesis de Maestría en Ingeniería en Sistemas de Información. UTN Buenos Aires.

[8] Straccia, L.; Pollo-Cattaneo, M.F.; Maulini, A. (2022). Visões da Gestão do Conhecimento para a construção de um modelo integral. En Proceeding in Congresso Brasileiro de Gestão do Conhecimento, São Paulo, Brasil.

[9] Straccia, L.; Maulini, A.; Ramacciotti, C.; Pollo-Cattaneo, M.F. Fases propuestas para el diseño y construcción de un modelo de Gestión del Conocimiento. En Serna, E. (2021). Desarrollo e Innovación en Ingeniería. Ed. Instituto Antioqueño de Inv., Medellín.

[10] Straccia, L.; Pollo-Cattaneo, M.F.; Maulini, A. Knowledge Management Model: A Process View. En ICCSA 2023. Lecture Notes in Computer Science, Springer, Cham.

[11] Straccia, L.; Ramacciotti, C.; Pollo-Cattaneo, M.F. Una visión de la tecnología para la Gestión del Conocimiento. Resultados en la literatura latinoamericana. En Serna, E. (2020). Desarrollo e Innovación en Ingeniería. Ed, Instituto Antioqueño de Inv., Medellín.

[12] Straccia, L.; Maulini, A.; Bongiorno, M.G.; Giorda, M.; Pollo-Cattaneo, M.F. Knowledge Representation and Technologies in the Latin American Academic Literature. ICAIW 2022: Workshops at ICAI, Arequipa, Perú.

[13] Straccia, L.; Pollo-Cattaneo, M.F.; Maulini, A. Knowledge management technologies for a n-layered architecture. ICAIW 2021: Workshops at ICAI, Buenos Aires, Argentina.

[14] Straccia, L.; Pollo-Cattaneo, M.F.; Giorda, M.; Bongiorno, M.G.; Maulini, A. Architecture on Knowledge Management Systems: its presence in the academic literature. International Conference ICAI 2022, Arequipa, Perú.

[15] Pérez López Portillo, H.: Gestión y Medición del Conocimiento en Organizaciones Públicas, Master's Thesis, Universidad de Guanajuato, México (2016).

[16] Straccia, L.; Pollo-Cattaneo, M.F. (2023). Descriptors for Technology Key Area in a Knowledge Management Maturity Model. ICAIW 2023: Workshops at ICAI, Guayaquil, Ecuador.

[17] Zapata Guerrero, E. Las PyMEs y su problemática empresarial. Análisis de casos. Revista Escuela de Administración de Negocios, 52, Universidad EAN. Bogotá, Colombia (2004)

[18] Lavarone, G.: Costos por órdenes de producción: su aplicación a la industria panificadora. Universidad Nacional de Cuyo, Mendoza (2012)

[19] Clemente Almendros, J. (dir): Estudio de la pequeña y mediana empresa. Estrategias para mejorar la competitividad de la PyME en la Región de La Rioja. Un análisis comparativo con España. Universidad Internacional de La Rioja (2023)

[20] Bohorquez, M., Martinez Palacio, U., Torres Royero, W. Guía metodológica de adquisición de software para pequeñas y medianas empresas del sector privado. Corporación Universitaria de la Costa, Barranquilla (2010).

[21] Ministerio de Economía de la Nación, Argentina. Disponible en https://www.argentina.gob.ar/noticias/las-pymes-argentinas-generaron-140000-puestos-de-trabajo-en-el-primer-trimestre

DESARROLLO DE HERRAMIENTAS PARA LA OPERACIONALIZACIÓN DE BUENAS PRÁCTICAS DE CALIDAD DE SOFTWARE

Emanuel Irrazabal; Gladys Dapozo; Cristina Greiner; Andrea Lezcano Airaldi; Juan Andrés Carruthers; María de los Ángeles Ferraro; Lourdes Romera

Facultad de Ciencias Exactas y Naturales y Agrimensura.
Universidad Nacional del Nordeste
{eirrazabal, alezcano}@exa.unne.edu.ar

RESUMEN

Esta línea se enmarca en el Proyecto F005-2021; una continuación de los proyectos F07-2009, F10-2013 y F018-2017 enfocados en el desarrollo de modelos, métodos y herramientas para la calidad del software teniendo en cuenta el entorno regional. Este proyecto aborda los temas emergentes en el área de la calidad de software, en particular, aspectos referidos a la gestión de proyectos en administraciones públicas regionales y el desarrollo de un modelo para la evaluación de visualizaciones con técnicas narrativas o de *storytelling*.

En particular, se está trabajando en la construcción de una herramienta para la evaluación de la calidad y mejora asistida de visualizaciones aplicando técnicas narrativas, en un modelo para sistemas de recomendación de requerimientos no funcionales en sistemas inteligentes y un conjunto de experimentos para comprobar el uso de ChatGPT en la elaboración de historias de usuario.

Palabras clave: calidad de software, visualización de la información, ChatGPT, historias de usuario

CONTEXTO

Las líneas de Investigación y Desarrollo presentada en este trabajo corresponden al proyecto PI-21F005 “Personalización de buenas prácticas de calidad de software en la región”, acreditado por la Secretaría de Ciencia y Técnica de la Universidad Nacional del Nordeste (UNNE) para el periodo 2022-2025, la beca interna doctoral de CONICET otorgada por RESOL-2021-154-APN-DIR#CONICET para el período 2021-2025.

1. INTRODUCCIÓN

Los procesos de desarrollo de software actual incluyen, de forma cada vez más frecuente, técnicas y algoritmos de aprendizaje automático (*machine learning*) para, por ejemplo, apoyar los procesos de toma de decisiones, o la automatización de tareas repetitivas, entre otras posibilidades [1]. En la literatura se pueden identificar diversas aplicaciones exitosas en múltiples dominios, como la gestión de préstamos, la decisión de contratación de empleados y el campo de la biología [2,3]. Sin embargo, a pesar de los beneficios que aporta el aprendizaje automático, se han planteado cuestiones referidas a las implicancias que dichas técnicas pueden tener en la ética e imparcialidad o equidad [1].

La imparcialidad se refiere a la capacidad de los sistemas para tomar decisiones no sesgadas [4]. El sesgo se define como un error sistemático en los procesos de toma de decisiones y puede surgir de diversas fuentes, incluyendo la recolección de datos, el diseño de algoritmos y la interpretación humana [5]. En este contexto, la dependencia de datos históricos puede llevar a un algoritmo de aprendizaje automático a adquirir conocimientos sesgados sobre las relaciones que rigen un fenómeno, lo que podría generar predicciones y recomendaciones injustas, perpetuando la discriminación y la injusticia. Por lo tanto,

la definición de métodos y herramientas capaces de reducir los riesgos relacionados con preocupaciones éticas representa un desafío clave [1].

Recientemente, se ha argumentado la necesidad de nuevos procesos de ingeniería para tratar los requerimientos relacionados a la ética a lo largo del ciclo de vida del software [1]. Sin embargo, operacionalizar esta necesidad resulta complejo debido a la naturaleza de dichos requerimientos y a la falta de regulaciones y estándares [3]. Por un lado, depende estrictamente del dominio de aplicación y de la tarea específica para la cual se diseña un sistema inteligente, por ejemplo, en relación con la imparcialidad, una característica puede ser sensible en un contexto y no en otro. Por otro lado, pueden existir múltiples definiciones que abordan diversas perspectivas [6], también dependientes del dominio de aplicación [7].

En la literatura es posible encontrar métodos para dar soporte al desarrollo de sistemas inteligentes con características de imparcialidad (por ejemplo, [1] y [8]). Sin embargo, la mayoría de estos métodos solo pueden usarse en etapas avanzadas del desarrollo, por ejemplo, durante el entrenamiento del sistema, mientras que, el desarrollo de técnicas que permitan a los profesionales realizar análisis de imparcialidad tempranos posibilitando el tratamiento de la imparcialidad a lo largo de todo el ciclo de vida del desarrollo ha recibido menos atención [1]. La gestión de requerimientos es una fase principal en el desarrollo de software, que involucra la identificación, documentación y mantenimiento de funcionalidades que el sistema debe cumplir, tanto para grandes fábricas de desarrollo software como para pequeños equipos de trabajo regionales [9][10]. En particular, los requerimientos no funcionales (RNFs) especifican características de calidad más allá de las funcionalidades básicas del sistema [11], como la operabilidad, la mantenibilidad, la seguridad o la usabilidad. Sin embargo, no es claro si el conocimiento acumulado sobre los RNFs sigue siendo aplicable para el desarrollo de sistemas inteligentes [12], necesitándose como en otros casos la construcción de metodologías de trabajo específicas que operacionalizan las buenas prácticas [13].

Finalmente, otra línea de investigación es acerca de la narrativa basada en datos, una estrategia efectiva para transmitir información, siendo su propósito estimular la atención y el compromiso (o *engagement*) de la audiencia y facilitar la toma de decisiones [14]. En el contexto actual, donde las posibilidades de toma de datos de distintos sistemas software ha tenido un crecimiento explosivo (Big Data), la visualización de datos proporciona un medio valioso por permitir un análisis asistido (por ejemplo, mediante controles de sumarización, división de los datos, búsquedas contextuales, o gráficos interactivos) [15]. Más aún, dado que muchas veces se utilizan para la toma de decisiones, es importante considerar las perspectivas de los usuarios finales en la construcción de dichas narrativas. Esta interrelación entre: i) las fuentes de datos que dan soporte a una visualización y ii) los intereses, preferencias y necesidades de los destinatarios de la misma, determina en gran medida la calidad de las narrativas a construir. No considerar estos aspectos suele ocasionar problemas relacionados con: baja calidad del contenido visual, pobre utilización de las fuentes de datos, ineficiencia en el desarrollo del contenido y, en general, poca entrega de valor.

En el contexto actual de desarrollos ágiles, uno de los problemas recurrentes es la poca calidad del contenido visual [16]. Por ello, la implementación de las buenas prácticas de *storytelling* en la construcción de un sistema y su posterior evaluación, pueden tener un impacto significativo en la calidad del producto resultante y en la entrega de valor por parte del sistema a los usuarios finales [17][18].

Se han realizado varios estudios que incluyen las prácticas y pautas generales a seguir para crear visualizaciones efectivas [19-21]. Si bien se han propuesto algunos

criterios para evaluaciones parciales, no se encuentran trabajos exhaustivos, herramientas que asistan a ello o propuestas sistemáticas de modelos para evaluar las narrativas basadas en datos.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Las líneas de investigación presentadas se centran en mejorar la calidad del software mediante el desarrollo de métodos y herramientas que aborden problemáticas actuales en visualización de datos, inteligencia artificial y aprendizaje automático. Un aspecto clave de esta línea durante el año 2025 será la identificación y mitigación de sesgos en los algoritmos de toma de decisiones. Dado que los sistemas basados en datos históricos pueden perpetuar discriminación e injusticia, se buscan estrategias para garantizar la imparcialidad en el desarrollo de software.

La propuesta incluye el diseño de mecanismos que reduzcan errores sistemáticos y favorezcan la equidad en los resultados generados por estos sistemas.

Otro eje de estudio se orienta a la incorporación de principios éticos en el ciclo de vida del software. A pesar del reconocimiento de la importancia de estos principios, su implementación efectiva sigue siendo un desafío debido a la falta de regulaciones y estándares definidos. La investigación se enfoca en la creación de marcos metodológicos que permitan incorporar consideraciones éticas desde las etapas iniciales del desarrollo, adaptándose a distintos dominios y contextos de aplicación. Esto incluye la exploración de definiciones de imparcialidad, estrategias de mitigación de sesgos y su aplicación en diferentes áreas del conocimiento.

Por último, se aborda la narrativa basada en datos como un elemento fundamental en la comunicación de información generada por sistemas inteligentes. En un contexto de creciente disponibilidad de datos, la investigación se enfoca en operacionalizar el modelo NaviQ, construido por el grupo de investigación para evaluar la calidad de las visualizaciones y narrativas digitales.

3. RESULTADOS OBTENIDOS/ESPERADOS

El objetivo principal de la investigación propuesta para el año 2025 es definir un marco de trabajo para la identificación de aspectos no funcionales relacionados al desarrollo ético de sistemas inteligentes. Principalmente, se estudiarán los atributos de imparcialidad y explicabilidad en dichos sistemas. Este objetivo se trabajará al inicio en las siguientes actividades específicas, atendiendo tanto aspectos metodológicos como tecnológicos:

1. Identificar y analizar los requerimientos relacionados al desarrollo ético de sistemas inteligentes describiendo técnicas para su abordaje.
2. Explorar e implementar modelos de aprendizaje automático para la identificación de aspectos relacionados al desarrollo ético en los requerimientos.

Narrativa digital

En el marco de esta línea de trabajo y teniendo en cuenta la estrategia de investigación Design Science [22] se llevaron adelante diferentes estudios de caso que han ido cimentando los resultados posteriores:

Estudio de caso [23] donde se determinaron los beneficios de incluir las buenas prácticas de visualización de datos en el desarrollo de un sistema realizado en el contexto de crisis sanitaria con tiempos límite restringidos. Para ello, se analizaron 16 gráficos de un sistema de información que gestiona el seguimiento del aislamiento y los permisos de circulación en la cuarentena por la pandemia de COVID-19.

Los resultados indican que el uso de técnicas de narrativa basada en datos contribuye a facilitar el proceso de toma de decisiones mediante la implementación de buenas prácticas de visualización que

aumentan la comprensión y la memorabilidad de los gráficos.

Estudio de caso [24] con el objetivo de desarrollar e implementar una aplicación web para un grupo de investigación inserto en una institución académica, utilizando una historia como hilo conductor para capturar y retener la atención de los usuarios.

Los resultados evidenciaron en general, una experiencia positiva al navegar el sitio para realizar las tareas propuestas. Esto podría deberse a que se planteó al usuario como el "protagonista" de la historia, apelando de esta manera a la interacción con el sistema para obtener información.

Una Revisión Sistemática de la Literatura presentada y aprobada ha caracterizado las buenas prácticas en la construcción de visualizaciones de datos y las estrategias de evaluación reportadas al momento. En ese contexto, se construyó el modelo de evaluación de visualizaciones narrativas NaviQ y la herramienta que la operacionaliza. Junto con ello se desarrollaron dos estudios de caso con empresas de la región nordeste argentino y con equipos de trabajo de la Universidad de Hamburgo en el marco de una estancia doctoral de investigación patrocinada por el Servicio Alemán de Intercambio Académico (DAAD) de Alemania.

4. FORMACIÓN DE RECURSOS HUMANOS

En esta línea de trabajo del Grupo de Investigación sobre Calidad de Software (GICS) están involucrados 4 docentes investigadores y 3 becarios internos doctorales CONICET.

5. REFERENCIAS

[1] C. Ferrara, F. Casillo, C. Gravino, A. De Lucia, and F. Palomba, "ReFAIR: Toward a Context-Aware Recommender for Fairness Requirements Engineering," in *Proceedings of the IEEE/ACM 46th International Conference on Software Engineering*, 2024, pp. 1–12, doi: 10.1145/3597503.3639185.

[2] J. Dastin, "Insight - Amazon scraps secret AI recruiting tool that showed bias against women," *Reuters*, 2018.

[3] L. Baresi, C. Criscuolo, and C. Ghezzi, "Understanding Fairness Requirements for ML-based Software," in *2023 IEEE 31st International Requirements Engineering Conference (RE)*, 2023, pp. 341–346, doi: 10.1109/RE57278.2023.00046.

[4] S. Verma and J. Rubin, "Fairness definitions explained," in *Proceedings of the International Workshop on Software Fairness*, 2018, pp. 1–7, doi: 10.1145/3194770.3194776.

[5] E. Ferrara, "Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies," *Sci*, vol. 6, no. 1, p. 3, Dec. 2023, doi: 10.3390/sci6010003.

[6] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, "A Survey on Bias and Fairness in Machine Learning," *ACM Comput. Surv.*, vol. 54, no. 6, pp. 1–35, Jul. 2022, doi: 10.1145/3457607.

[7] J. Chakraborty, S. Majumder, and T. Menzies, "Bias in machine learning software: why? how? what to do?," in *Proceedings of the 29th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, 2021, pp. 429–440, doi: 10.1145/3468264.3468537.

[8] R. Brower-Sinning, G. A. Lewis, S. Echeverría, and I. Ozkaya, "Using Quality Attribute Scenarios for ML Model Test Case Generation," Jun. 2024.

[9] Dapozo, Gladys N., et al. "Características del desarrollo de software en la ciudad de Corrientes." XXI Congreso Argentino de Ciencias de la Computación (Junín, 2015). 2015.

[10] Coimbra, Rodrigo Pereyra, et al. "Personalización de técnicas ágiles en el desarrollo de software para la obtención de requisitos de calidad en Pymes: una revisión sistemática de la literatura." SADIO Electronic Journal of Informatics and Operations Research 19.2 (2020): 76-93.

[11] E. Soremekun, M. Papadakis, M. Cordy, and Y. Le Traon, "Software Fairness: An Analysis and Survey," May 2022.

[12] L. Chung, B. A. Nixon, E. Yu, and J. Mylopoulos, *Non-functional requirements in software engineering*. 2012.

[13] García, M., Emanuel Irrazabal, and Javier Garzás. "Implantación de las normas ISO/IEC 15504 e ISO/IEC 12207 con métodos ágiles y SCRUM." Agil Spain (2010).

[14] Nussbaumer Knaflic, C. Storytelling with Data: A data visualization guide for business professionals. Hoboken, New Jersey: John Wiley & Sons, Ltd., 2015.

[15] Kosara, R., Mackinlay, J. Storytelling: The Next Step for Visualization. Computer, Vol 46, Issue 5, 44-50 2013

[16] Nicoletti, M., Schiaffino, S., Díaz Pace, J. A.: An optimization-based tool to support the cost-effective production of software architecture documentation. J. Softw. Evol. Process. 27(9): 674-699 (2015).

[17] Baker, R. Agile UX Storytelling. 2017.

[18] Zimmerman, B. Applying Tufte's principles of information design to creating effective Web sites. ACM SIGDOC Annu. Int. Conf. Comput. Doc. Proc., pp. 309–317, 1997, doi: 10.1145/263367.263406.

[19] Kosara, R. An Argument Structure for Data Stories. Proc. Eurographics/IEEE VGTC Symposium on Visualization (EuroVis), 2017.

[20] Tufte, E. R.The Visual Display of Quantitative Information. 1983.

[21] Ram D., et al. Case Report: Rapid Development of Visualization Dashboards to Enhance Situation Awareness of COVID-19 Telehealth Initiatives at a Multi-Hospital Healthcare System. Journal of the American Medical Informatics Association (2020).

[22] Johannesson, Paul, and Erik Perjons. An introduction to design science. Vol. 10. Cham: Springer, 2014.

[23] Lezcano Airaldi A., Diaz-Pace J.A., Irrazábal, E. (2021) Data-driven Storytelling to Support Decision Making in Crisis Settings: A Case Study. JUCS - Journal of Universal Computer Science 27(10): 1046-1068. https://doi.org/10.3897/jucs.66714

[24] Lezcano Airaldi, A., Sandoval, D. & Irrazábal, E. (2021). Storytelling Aplicado al Diseño de Sitios Web: Un Estudio de Caso. En IV Jornadas de Calidad de Software y Agilidad (pp. 37-46).

Enfoques y Tendencias en el Desarrollo de Aplicaciones para Dispositivos Móviles.

Pablo Thomas, Leonardo Corbalán, Juan Fernández Sosa, Fernando Tesone, Verena Olsowy, Patricia Pesado

Instituto de Investigación en Informática LIDI (III-LIDI)

Facultad de Informática – Universidad Nacional de La Plata

50 y 120 La Plata Buenos Aires

Centro Asociado a la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC)

526 e/ 10 y 11 La Plata Buenos Aires

{pthomas, corbalan, jfernandez, ftesone, volsowy, ppesado}@lidi.info.unlp.edu.ar

RESUMEN

Esta línea de investigación se centra en el estudio de la Ingeniería de Software aplicada a las diversas tendencias en el desarrollo de aplicaciones móviles resilientes, incluyendo la utilización de tecnologías como *blockchain* y su posible aplicación en el contexto de las ciudades inteligentes.

Palabras claves: Plataformas para Dispositivos Móviles – Aplicaciones Nativas – Aplicaciones Híbridas – Aplicaciones Interpretadas – Aplicaciones por Compilación Cruzada – Aplicaciones Web Progresivas – Offline First – Instant App – Resiliencia – Ciudades Inteligentes – Blockchain.

CONTEXTO

Esta línea de Investigación forma parte del Proyecto "*Diseño, desarrollo y evaluación de sistemas en escenarios híbridos para áreas clave de la sociedad actual: educación, ciudades inteligentes y gobernanza digital*" del Instituto de Investigación en Informática LIDI de la Facultad de Informática, acreditado por el Ministerio de Educación de la Nación.

Hay cooperación con Universidades de Argentina y se está trabajando con Universidades de Europa y Latinoamérica.

1. INTRODUCCIÓN

Los avances en tecnologías de comunicación y computación han propiciado el surgimiento de la 'Computación móvil' [1], la cual enfrenta importantes desafíos en el ámbito del desarrollo de software. Entre estos desafíos se incluyen la gestión de diversos estándares, protocolos y plataformas tanto de hardware como de software, además de las limitaciones propias de los dispositivos móviles, tales como el procesamiento, el consumo energético y el almacenamiento.

La computación móvil ha impulsado el desarrollo de nuevas metodologías y prácticas que han contribuido al fortalecimiento de la Ingeniería de Software como disciplina, acompañando la evolución tecnológica en este ámbito.

El crecimiento del mercado de aplicaciones móviles se ha visto impulsado por el aumento de dispositivos y la migración de empresas hacia servicios basados en la web. Para mantenerse competitivas, las compañías deben reducir los tiempos de desarrollo y asegurar la compatibilidad con la mayor cantidad de dispositivos posible. No obstante, la fragmentación del ecosistema, derivada de la diversidad de hardware, sistemas operativos y plataformas de desarrollo, representa un desafío significativo.

Una forma de afrontar esta problemática es optar por el desarrollo nativo para cada plataforma de destino, utilizando las herramientas, tecnologías y lenguajes específicos de cada una.

El desarrollo nativo de aplicaciones móviles presenta diversas ventajas, como el acceso total a las funcionalidades del dispositivo, un alto rendimiento y una interfaz gráfica adaptada a la plataforma. También permite operar sin conexión a Internet, ejecutar procesos en segundo plano y distribuir las aplicaciones a través de tiendas en línea. Sin embargo, esta estrategia implica un costo

elevado, ya que el código fuente no es reutilizable entre plataformas, lo que incrementa el esfuerzo y los costos asociados al desarrollo, mantenimiento y distribución de nuevas versiones.

Como alternativa al desarrollo nativo, los enfoques multiplataforma permiten reutilizar el código fuente para crear versiones compatibles con distintas plataformas [2]. Entre ellos, las aplicaciones Web Móviles, desarrolladas con tecnologías como HTML, CSS y JavaScript, operan dentro de navegadores sin necesidad de ajustarse a un sistema operativo móvil en particular. Sin embargo, presentan limitaciones en cuanto a rendimiento y seguridad, además de ofrecer una experiencia de usuario generalmente inferior a la de las aplicaciones nativas.

Dentro de esta categoría, un concepto innovador es el de las Aplicaciones Web Progresivas (PWA), las cuales aprovechan tecnologías web avanzadas para proporcionar experiencias similares a las aplicaciones nativas. Éstas buscan garantizar un alto rendimiento, tiempos de carga reducidos y la posibilidad de operar sin conexión a Internet (*offline first*).

Para facilitar la integración de las PWAs en el ecosistema Android, Google ha desarrollado la tecnología TWA, que permite ejecutarlas en dispositivos móviles a través de Google Play Store sin mostrar la interfaz del navegador.

Otra estrategia multiplataforma es el desarrollo híbrido, que emplea tecnologías web estándar dentro de un contenedor especial que facilita el acceso a las funcionalidades del dispositivo mediante una API específica. Aunque estas aplicaciones pueden ser distribuidas en tiendas oficiales, la inclusión de componentes no nativos puede afectar el rendimiento y la fluidez de la interfaz, debido a la sobrecarga generada por el contenedor web.

El enfoque interpretado para el desarrollo de aplicaciones móviles permite generar aplicaciones que, en su mayoría, son compiladas a código nativo, con ciertos elementos interpretados en tiempo de ejecución. Estas soluciones se implementan sin depender directamente de una plataforma en particular, utilizando tecnologías y lenguajes como JavaScript, TypeScript y XML. Aunque este enfoque permite interfaces nativas, la creación de nuevos componentes puede ser compleja debido a los desafíos de compatibilidad entre múltiples plataformas.

Finalmente, otro método para el desarrollo multiplataforma es la compilación cruzada, que permite generar versiones nativas de alto rendimiento para cada plataforma de destino a partir de un único código fuente.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

- Enfoques de desarrollo de Aplicaciones para Dispositivos Móviles.
- Metodologías y Técnicas de la Ingeniería de Software y su aplicación en el desarrollo de software para dispositivos móviles.
- Aplicaciones Nativas en Android [3].
- Aplicaciones Nativas en iOS [3].
- Aplicaciones Web Móviles.
- Aplicaciones Móviles Híbridas (PhoneGap [4], Ionic [5]).
- Aplicaciones Móviles Interpretadas (Appcelerator Titanium [6], NativeScript [7]).
- Aplicaciones Móviles generadas por compilación cruzada (Xamarin [8], Corona [9]).
- PWA, Instant App y Offline First.
- TWA (Trusted Web Activities),
- Análisis y estudio comparativo de requerimientos no funcionales, tales como rendimiento, consumo de energía, tamaño de software, entre otros, en los distintos enfoques de Aplicaciones Móviles.
- Experiencia de usuario en Aplicaciones Móviles generadas con distintos enfoques de desarrollo.

- El rol de los dispositivos móviles en aplicaciones de *crowdsourcing* o colaboración abierta distribuida.
- El rol de las aplicaciones móviles como herramientas para hacer frente a la pandemia de COVID-19.
- Aplicaciones Móviles resilientes. Escenarios adversos: conectividad inestable, pérdida de conexión, ralentización del tráfico de red, ciber-ataques. Estrategias: *offline first*, almacenamiento local, memoria cache, *retry pattern* (reintento) *throttling pattern* (estrangulamiento), lineamientos para el desarrollo de apps móviles resistentes a ciber-ataques.
- Ciudades inteligentes: El rol de las aplicaciones móviles y de la tecnología *blockchain*.
- Internet de las Cosas (IoT), Internet de Todo (IoE), dispositivos beacons y aplicaciones de sensado móvil o sensado urbano.
- Computación Cuántica. Herramientas educativas para facilitar el aprendizaje de la programación cuántica y su aplicación en diversos ámbitos. Análisis del impacto de los avances en esta tecnología en el desarrollo de ciudades inteligentes.

3. RESULTADOS OBTENIDOS/ESPERADOS

Los resultados obtenidos/esperados se pueden resumir en:

- Se ha evaluado el impacto del enfoque de desarrollo en el rendimiento de las aplicaciones, considerando plataformas como iOS y Android, junto con diversos *frameworks* multiplataforma. Los resultados y conclusiones se presentaron en [10] y [11].
- Se creó la aplicación móvil "*Informática UNLP*" [12] para mejorar la comunicación entre alumnos y docentes de la Facultad de Informática de la Universidad Nacional de La Plata, incorporando una pizarra virtual adaptable y tecnología de realidad aumentada para mostrar la ocupación de aulas en tiempo real. Está disponible en iPhone, iPad y Android.
- En [13] se examinaron las implicaciones de la elección del enfoque de desarrollo en la eficiencia energética de las aplicaciones, abarcando aquellas con alta carga de procesamiento, reproducción de video y acceso a imágenes.
- En [14] se publicaron los resultados de un estudio sobre el impacto de 9 tecnologías de desarrollo de aplicaciones móviles sobre 23 características relevantes para la Ingeniería de Software.
- En [15] se investigó cómo diferentes *frameworks* de desarrollo multiplataforma afectan el tamaño de las aplicaciones construidas. Los resultados son de interés dado que el espacio disponible se ha vuelto crítico para muchos usuarios.
- Se cuantificó el impacto de la elección del *framework* de desarrollo en tres requerimientos no funcionales prioritarios para los usuarios de dispositivos móviles: rendimiento, consumo de energía y uso de almacenamiento. Los resultados y conclusiones se presentaron en [16], ofreciendo una guía para desarrolladores que requieren seleccionar el *framework* más adecuado.
- Se analizaron las Aplicaciones Web Progresivas (PWA) en [17], comparándolas con las aplicaciones web móviles y destacando diferencias y similitudes, identificando ventajas y desventajas de ambos enfoques.
- Se desarrolló "*InnovApp*", una PWA que permite una visita virtual guiada a la muestra de Ciencia y Tecnología 2019 en el Centro de Innovación y Transferencia Tecnológica (CIyTT) de la Facultad de Informática de la UNLP
- Se llevó a cabo un estudio sobre el uso de la tecnología móvil para enfrentar la pandemia de COVID-19, analizando las aplicaciones promovidas por los gobiernos



Figura 1. Imagen alusiva. Aplicaciones Móviles, Blockchain, IOT. Ciudades inteligentes

de 22 países como herramientas tecnológicas para combatir la crisis. Los resultados y conclusiones se presentaron en [18].

- Se implementó un prototipo de sistema de sensado móvil participativo para que los ciudadanos monitoreen la calidad del agua de red en sus hogares y alerten sobre anomalías a la comunidad y autoridades competentes [19].
- Se elaboraron recomendaciones para mejorar la resiliencia de las aplicaciones móviles ante un aumento repentino de la carga computacional del dispositivo y el consumo de energía [20].
- En [21], se exploraron las oportunidades de interacción entre dispositivos móviles y *beacons*, pequeños dispositivos inalámbricos de tecnología BLE que se utilizan en IoT y Ciudades Inteligentes para aplicaciones como marketing, servicios de proximidad y seguimiento de objetos.
- En [22] se revisaron las principales herramientas para la programación cuántica, enfatizando su papel clave en la democratización de esta tecnología y su integración educativa, considerando aspectos como facilidad de aprendizaje, robustez, disponibilidad y aplicaciones formativas.
- Actualmente, la investigación continúa explorando estrategias para aumentar la resiliencia de las aplicaciones móviles en diversas circunstancias adversas.
- Se continúa desarrollando "*Informática UNLP*", una aplicación multiplataforma en constante evolución
- Se prosigue con el estudio y el análisis del alcance de la tecnología de sensado móvil en aplicaciones de Internet de las Cosas (IoT), Internet de Todo (IoE) y sensado urbano.
- Se profundizará sobre diversas formas de persistencia de datos en aplicaciones móviles según el enfoque/*framework* de desarrollo utilizado como estrategia para resistir la inestabilidad de las conexiones.
- Se sigue investigando el papel de las Aplicaciones Móviles en ciudades inteligentes y el potencial de la tecnología *blockchain* para mejorar la calidad de vida urbana. Este estudio analizará cómo la

blockchain puede ofrecer soluciones innovadoras en áreas como transparencia y trazabilidad de datos, gestión de residuos, seguimiento de energía y agua, y creación de mercados energéticos descentralizados, entre otros.

- Se continuará con el estudio de herramientas para la programación cuántica, analizando su evolución y su impacto en la accesibilidad y eficiencia del desarrollo de software cuántico. Además, se estudiará el impacto de la computación cuántica en la conformación de las ciudades inteligentes del futuro, explorando cómo esta tecnología puede optimizar la gestión de datos, la seguridad de la información y la toma de decisiones en entornos urbanos cada vez más interconectados.
- Se Promoverá el avance sostenido y continuo de la formación de los miembros involucrados en esta línea de investigación.

4. FORMACIÓN DE RECURSOS HUMANOS

Los integrantes de esta línea de investigación dirigen Tesinas de Grado y Tesis de Postgrado en la Facultad de Informática, y Becarios III-LIDI en temas relacionados con este proyecto. Además, participan en el dictado de asignaturas/cursos de grado y postgrado de la Facultad de Informática de la UNLP.

5. BIBLIOGRAFÍA

1. Hongxing Li, Guochu Shou, Yihong Hu, Zhigang Guo. Mobile Edge Computing: Progress and Challenges. 2016 4th IEEE International Conference on Mobile Cloud Computing, Services, and Engineering (MobileCloud). Oxford UK.

2. Spyros Xanthopoulos, Stelios Xinogalos, A Comparative Analysis of Cross-platform Development Approaches for Mobile Applications, BCI' 2013, Greece

3. Tracy, K.W., Mobile Application Development Experiences on Apple's iOS and Android OS, Potentials, IEEE, 2012.

4. http://phonegap.com/

5. https://ionicframework.com/

6. http://www.appcelerator.com/

7. https://www.nativescript.org/

8. http://xamarin.com/

9. https://coronalabs.com/

10. Delia L., Galdamez N., Corbalan L., Thomas P, Pesado P., Un Análisis comparativo de rendimiento en Aplicaciones Móviles Multiplataforma, CACIC 2015, UNNOBA Junín, Octubre 2015.

11. Delia L., Galdamez N., Corbalan L., Thomas P., Pesado P. Approaches to Mobile Application Development: Comparative Performance Analysis SAI Computing Conference (SAI), 2017. Londres, del 18 al 20 de Julio de 2017.

12. Fernández Sosa J., Cuitiño A., Thomas P., Delia L., Caseres G., Corbalán L., Pesado P. "Informática UNLP" la App de la Facultad de Informática. CACIC 2017. La Plata, del 9 al 13 de Octubre de 2017.

13. L. Corbalan, J. Fernandez Sosa, A. Cuitiño, L. Delía, G. Cáseres, P. Thomas, P. Pesado. "Development Frameworks for Mobile Devices: A Comparative Study about Energy Consumption". 5th IEEE/ACM International Conference on Mobile Software Engineering and Systems. Gothenburg, Sweden. 27 y 28 de Mayo del 2018.

14. L. Delia, P. Thomas, L. Corbalan, J. Fernandez Sosa, A. Cuitiño, G. Cáseres, P. Pesado. "Development Approaches for Mobile Applications: Comparative Analysis of Features" SAI - Computing Conference 2018. London, United Kingdom. 10 al 12 de Julio de 2018.

15. J. Fernández Sosa, P. Thomas, L. Delía, G. Cáseres, L. Corbalán, F. Tesone, A. Cuitiño, P. Pesado. "Mobile Application Development Approaches: A Comparative Analysis on the Use of Storage Space". XV Workshop de Ingeniería de Software - XXIV Congreso Argentino de Ciencias de la Computación

CACIC 2018. Tandil, Argentina. 8 al 12 de octubre de 2018.

16. Corbalán, L., Thomas, P., Delía, L., Cáseres, G., Sosa, J. F., Tesone, F., & Pesado, P. (2019, June). A Study of Non-Functional Requirements in Apps for Mobile Devices. In Conference on Cloud Computing and Big Data (pp. 125-136). Springer, Cham

17. Aguirre, V., Ortu, A., Delía, L. N., Thomas, P. J., Corbalán, L. C., Cáseres, G., & Pesado, P. M. (2019). PWA para unificar el desarrollo Desktop, Web y Mobile. In XXV Congreso Argentino de Ciencias de la Computación (CACIC) (Universidad Nacional de Río Cuarto, Córdoba, 14 al 18 de octubre de 2019).

18. Fernandez Sosa, J., Aguirre, V., Delía, L., Thomas, P., Corbalán, L., & Pesado, P. M. (2020). COVID-19: un análisis comparativo de Apps. In XXVI Congreso Argentino de Ciencias de la Computación (CACIC) (Universidad Nacional de La Matanza, 5 al 9 de octubre de 2020).

19. Fernández Sosa, J. F., Aguirre, V., Corbalán, L. C., Delía, L. N., Thomas, P. J., & Pesado, P. M. (2021). Sensado móvil como estrategia de participación ciudadana en Ciudades Inteligentes. In XXVII Congreso Argentino de Ciencias de la Computación (CACIC) (Modalidad virtual, 4 al 8 de octubre de 2021).

20. Corbalán, L. C., Fernández Sosa, J. F., Tesone, F., Dapoto, S. H., Cristina, F., Thomas, P. J., & Pesado, P. M. (2022). Recommendations for Resilient App Development. In X Jornadas de Cloud Computing, Big Data & Emerging Topics (La Plata, 2022).

21. Fernández Sosa, J. F., Medina, S., Thomas, P. J., Corbalán, L. C. & Naiouf, M. Análisis comparativo de alternativas para la detección y comunicación con dispositivos BLE Beacons en Aplicaciones Móviles Multiplataforma. En XXIX Congreso Argentino de Ciencias de la Computación (CACIC 2023). (Universidad de Luján, 9 al 12 de octubre de 2023).

22. Marrero. L., Olsowy. V., Fernandez Sosa, J., Tesone. F., Corbalan, L., Thomas, P., y Pesado, P. Una Revisión de Herramientas para Programación Cuántica, En XXX Congreso Argentino de Ciencias de la Computación (CACIC 2024). Facultad de Informática, La Plata, Octubre 2024. ISBN 978-950-34-2428-5

Especificación de Requisitos de Calidad de Software en Historias de Usuario: Una propuesta aplicando IA generativa como soporte

Carrizo Claudio*, Javier Saldarini*, Juan Pablo Mansilla*, Carlos Salgado+, Alberto Sánchez+, Mario Peralta+
*Facultad Regional San Francisco – Universidad Tecnológica Nacional
Av. de la Universidad 501 - San Francisco - Córdoba - Tel. 03564-421147
{cjcarrizo77, saldarinijavier, juampimansilla17}@gmail.com
+ Departamento de Informática - Facultad de Ciencias Físico-Matemáticas y Naturales Universidad Nacional de San Luis
Ejército de los Andes 950 – C.P. 5700 – San Luis – Argentina
{csalgado, alfanego, mperalta}@unsl.edu.ar

RESUMEN

El desarrollo de software es un proceso en constante transformación que comprende diversas etapas: el diseño, la implementación, las pruebas y el mantenimiento de aplicaciones y sistemas informáticos. A lo largo del tiempo, la evolución de la informática ha impulsado mejoras en las herramientas y tecnologías utilizadas en este ámbito, desde la adopción de lenguajes de programación de alto nivel, hasta la implementación de metodologías ágiles.

La evolución de la Inteligencia Artificial (IA) ha emergido también como un elemento que pretende ser clave en la optimización de los procesos de desarrollo de software. Herramientas basadas en IA, como asistentes de programación y modelos de aprendizaje automático, facilitan la generación de código, la detección de errores y la automatización de pruebas, reduciendo significativamente los tiempos de desarrollo y mejorando la precisión del software.

En ese contexto, la presente línea de investigación tiene como objetivo desarrollar una propuesta basada en IA generativa, que pueda ser utilizada como copiloto para dar soporte a la especificación de Requisitos de Calidad (RC) del Software, en contexto de desarrollo ágil de software, y utilizando el estándar de calidad ISO/IEC 25000 como marco de referencia.

Palabras clave: inteligencia artificial generativa, requisitos de calidad del software, ISO/IEC 25000, metodologías ágiles.

CONTEXTO

La presente línea de investigación se enmarca dentro del Proyecto de Investigación y Desarrollo UTN que lleva por título: *"Mejora de la calidad de los sistemas de información en entorno de desarrollo ágil"*.

El mencionado proyecto fue elaborado y está siendo ejecutado de manera conjunta entre los integrantes del Laboratorio de Calidad e Ingeniería de Software (LaCIS) – Facultad de Ciencias Físico-Matemáticas y Naturales, Universidad Nacional de San Luis, y los integrantes del Grupo de I+D denominado "Calidad del Software", perteneciente a la Universidad Tecnológica Nacional, Facultad Regional San Francisco.

El proyecto se encuentra homologado y financiado por la Secretaría de Ciencia, Tecnología y Posgrado de la Universidad Tecnológica Nacional, bajo el código EC9837 y según la Disp. SCTyP Nº 1-2023.

1. INTRODUCCIÓN

Una de las tendencias más recientes en la industria del software es la aplicación de la IA y el Aprendizaje Automático (AA). Estos avances han abierto nuevas posibilidades

para mejorar la eficiencia, la calidad y la colaboración en el desarrollo del software (Ocaña, Naranjo, & Fausto, 2023).

La IA Generativa se define como el sub-campo de la Inteligencia Artificial, que se centra en el desarrollo de modelos que pueden crear nuevos datos, como texto, imágenes, audio y video (Tsoftglobal, 2024). En particular, modelos de lenguaje como ChatGPT de OpenAI (openai, s.f.) y GitHub Copilot (github, s.f.), han demostrado un impacto considerable en la forma en que se desarrolla y mantiene el software (Ocaña, Naranjo, & Fausto, 2023).

Según expresa (Sommerville, 2011), el software se desarrolla de manera rápida para aprovechar las actuales oportunidades, con la finalidad de responder ante la amenaza competitiva, con lo cual, la entrega y el desarrollo rápido, son por lo general el requerimiento fundamental. En este sentido surgen y toman cada vez más preponderancia las metodologías ágiles para el desarrollo de software, las cuales se basan en el Manifiesto Ágil (Manifesto for Agile Software Development, s.f.), y son ampliamente adoptadas por la industria del software en la actualidad. Autores como Pressman (Pressman, 2005) y Sommerville (Sommerville, 2011) coinciden en que las más reconocidas en la industria del software son: Scrum (Scrum Alliance, s.f.) y Programación Extrema (Extreme Programming, s.f.).

Por otra parte, en el desarrollo ágil de software se promueve una gestión mínima de documentación que aporte valor a un usuario; a menudo se priorizan requisitos funcionales sobre los RC. Estos permiten describir la calidad deseada del software, y juegan un papel importante en el éxito de los proyectos de software; a menudo en el desarrollo ágil de software, los RC están mal definidos y también existe un conocimiento sintetizado o limitado sobre su manejo (Woubshet, y otros, 2020).

Somerville (Sommerville, 2011) expresa que un problema común con los requerimientos no funcionales o de calidad es que los usuarios o clientes, con frecuencia, proponen estos requerimientos como metas generales, donde esas metas suelen ser buenas intenciones, y dejan espacio para la interpretación y la disputa posterior, una vez que se entregue el sistema; también remarca, que siempre que sea posible, se deberán escribir de manera cuantitativa los requerimientos no funcionales, de manera que puedan ponerse objetivamente a prueba.

En ISO/IEC 25010 (ISO/IEC 25010:2011, 2011), se la define la calidad del software como “el grado en que el producto software satisface las necesidades expresadas o implícitas, cuando es usado bajo condiciones determinadas”.

En la familia de Normas ISO/IEC 25000 (ISO/IEC 25000:2014, 2014) encontramos un esquema organizado y orientado tanto a la especificación, como a la evaluación de RC, para garantizar la calidad de los productos software.

Según lo descripto anteriormente, se puede observar que actualmente, tanto las metodologías ágiles como la IA están teniendo un papel crucial y preponderante en la industria del software, lo cual, presupone una mayor productividad, eficiencia y calidad para los productos software desarrollados.

Es este contexto, se pone de manifiesto la necesidad de abordar la problemática de la especificación de los RC con una mirada puesta en tres aspectos: 1) las metodologías ágiles y sus artefactos para especificar requisitos; 2) la IA generativa como copiloto y soporte en las actividades para la especificación de RC de software; 3) las normas internacionales de referencia en calidad de producto software ISO/IEC 25000 (ISO/IEC 25000:2014, 2014).

El desarrollo de esta línea de investigación pretende llevar a cabo un estudio sobre los distintos modelos de IA generativos, y su aplicabilidad como agentes copiloto que

colaboren para especificar RC en el desarrollo de software guiado por metodologías ágiles, y tomando como marco de referencia lo propuesto por ISO/IEC 25000 (ISO/IEC 25000:2014, 2014).

Con el objetivo de arribar a una especificación de RC de Software partiendo de las Historias de Usuarios, y tomando como marco de referencia lo establecido en los estándares de calidad ISO/IEC 25010 (ISO/IEC 25010:2011, 2011) e ISO/IEC 25023 (ISO/IEC 25023:2016, 2016 (E)), es que se desarrollará una guía siguiendo el proceso establecido por ISO/IEC 25030 (ISO/IEC 25030:2007, 2007), la cual podrá ser utilizada por Analistas QA como input, para redactar los Prompt de un agente de IA Generativa como copiloto.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

La presente línea de investigación se compone de un conjunto de ejes de trabajo, los cuales se detallan a continuación:

1- **Elaboración del Estado del arte:** elaborar un estado del arte sobre la aplicabilidad de IA Generativa en los procesos de especificación de requisitos de calidad del software, en entornos de desarrollos ágiles, y utilizando estándares de calidad aplicados al software.

2- **Desarrollo de una Guía para redactar Prompt:** desarrollar una guía basada en el proceso de Análisis y Definición de RC del software propuesto por ISO/IEC 25030 (ISO/IEC 25030:2007, 2007), la cual esté compuesta por un conjunto de pasos, donde cada uno de ellos contendrá un guion para ser utilizado como Prompt, el cual permitirá interactuar con un agente de IA Generativa. Para desarrollar la mencionada Guía, se utilizará: un modelo para historias de usuario y criterios de aceptación; el modelo de calidad de producto definido por ISO/IEC 25010 (ISO/IEC 25010:2011, 2011); y las métricas de calidad, las cuales se obtendrán a través de la norma ISO/IEC 25023 (ISO/IEC 25023:2016, 2016 (E))

3- **Selección de un Agente de IA:** en función de los resultados del Eje 1, se deberán seleccionar al menos 3 agentes de IA Generativa, los cuales puedan actuar como Copiloto, para la aplicación de la guía desarrollada en el Eje 2. La precisión y coherencia que arrojen los resultados obtenidos por cada Copiloto de IA en cada paso de la guía, se compararán en función de trazabilidad con el estándar ISO/IEC 25000 (ISO/IEC 25000:2014, 2014), la historia de usuario de partida y sus criterios de aceptación.

4- **Ajustes de la Guía:** con la experiencia obtenida en el Eje 3, y habiendo analizado los resultados, se ajustará la Guía desarrollada en el Eje 2.

5- **Prueba de Concepto:** se pretende instanciar la Guía desarrollada en un caso real, a través del uso del Copiloto de IA Generativa seleccionado como resultado de las actividades llevadas a cabo en el Eje 3. Para ello, trabajaremos en conjunto con un Analista QA, sobre un proyecto real basado en el desarrollo de un sistema de gestión de emergencias y urgencias médicas. Como este proyecto se encuentra en desarrollo, se partirá de un conjunto de historias usuarios que ya hayan sido desarrolladas e implementadas, esto nos dará los puntos de comparación sobre las especificaciones de RC que se documentaron originalmente, versus las arrojadas con la implementación de la Guía, haciendo uso de un Copiloto de IA generativa.

6- **Análisis, Discusión y Publicación de los resultados obtenidos:** finalmente se llevará a cabo el análisis y discusión de los resultados; se pretende también publicar al menos un trabajo en una revista o congreso de la especialidad.

3. RESULTADOS

Si bien la línea de investigación que se presenta en este artículo es de reciente iniciación, el Proyecto de I+D que la contiene lleva dos años de ejecución, donde los resultados fueron publicados en distintos congresos y revistas de la especialidad. Los antecedentes mencionados forman parte del punto de partida para la línea de investigación aquí presentada.

Resultados Obtenidos

1- Se elaboró un estado del arte sobre la aplicabilidad de IA generativa en los procesos de especificación de requisitos de calidad del software, en entornos de desarrollos ágiles de software, y utilizando estándares de calidad del producto del software. De manera resumida se recopilaron antecedentes de trabajos académicos, sitios web académicos, y herramientas, que dan cuenta de la aplicación de IA Generativa en el campo del aseguramiento de la calidad del software en entornos de desarrollos ágiles.

2- Se desarrolló una Guía compuesta de 6 pasos (Ver Figura 1), siguiendo el proceso de Análisis y Definición de RC del software propuesto por ISO/IEC 25030 (ISO/IEC 25030:2007, 2007). En cada paso se genera un guion para ser utilizado como un Prompt, para poder interactuar con un agente de IA Generativa. La mencionada Guía utiliza un modelo para historias de usuario y criterios de aceptación, el modelo de calidad de producto definido por ISO/IEC 25010 (ISO/IEC 25010:2011, 2011) y las métricas de calidad de ISO/IEC 25023.



Figura 1. Diagrama de flujo de uso de la guía

Resultados Esperados

1- Informe que contenga la comparación respecto de la precisión y coherencia de los resultados obtenidos en cada paso de la aplicación de la Guía para cada Copiloto de IA; la comparativa estará dada en función de la trazabilidad con el estándar ISO/IEC 25000 (ISO/IEC 25000:2014, 2014), la historia usuario y sus criterios de aceptación.

2- Informe que contenga los resultados de la Prueba de Concepto realizada, a través de los puntos de comparación sobre las especificaciones de RC que se documentaron originalmente en las historias de usuario, versus las arrojadas con la implementación de la Guía y utilizando un Copiloto de IA generativa.

FORMACIÓN DE RECURSOS HUMANOS

La presente línea de investigación se trabaja desde el Grupo I+D "Calidad de Software", en colaboración con docentes

investigadores formados de la Universidad Nacional de San Luis. El equipo de trabajo se encuentra conformado 5 docentes investigadores formados, 2 investigadores en formación, 2 graduados y 2 estudiantes de grado avanzado.

En el contexto de la ejecución del proyecto, en donde se inserta la presente línea de investigación, se está desarrollando una tesis de "Maestría en Calidad del Software", y un trabajo final de carrera de un estudiante de Ingeniería en Sistemas de Información

5. BIBLIOGRAFÍA

Extreme Programming. (s.f.). Obtenido dehttps://www.agilealliance.org/glossary/xp

github. (s.f.). github. Obtenido de https://github.com/features/copilot

ISO/IEC 25000:2014. (2014). Systems and software engineering-Systems and software Quality Requirements and Evaluation (SQuaRE).

ISO/IEC 25010:2011. (2011). Systems and software engineering-Systems and software Quality Requirements and Evaluation (SQuaRE)-System and software quality models. ISO.

ISO/IEC 25023:2016. (2016 (E)). Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Measurement of system and software product quality.

ISO/IEC 25030:2007. (2007). Software engineering — Software product Quality Requirements and Evaluation (SQuaRE) — Quality requirements. ISO.

Manifesto for Agile Software Development. (s.f.). Obtenido de http://agilemanifesto.org/

Ocaña, L., Naranjo, F., & Fausto, A. (2023). Estudio neutrosófico sobre la afectación de la. Neutrosophic Computing and Machine Learning, 296-304.

openai. (s.f.). openai. Obtenido de https://openai.com/chatgpt/

Pressman, R. (2005). Ingeniería de Software. (6 ed.). Mcgraw-Hill.

Scrum Alliance. (s.f.). Obtenido de https://resources.scrumalliance.org/Article/quick-guide-things-scrum

Sommerville, I. (2011). Ingeniería de Software (9° ed.). México: PEARSON EDUCACIÓN.

Tsoftglobal. (09 de 2024). Tsoftglobal. Obtenido de https://lc.cx/MxaAFU

Woubshet, B., Pertti, K., Lidia, L., Xavier, B., Silverio, M.-F., Vollmer, A. M., . . . Oivo, M. (2020). Management of quality requirements in agile and rapid software develop,ment: A systematic mapping study. Information and Software Technology, 123, 106-225.

Explorando el *Diseño Participativo* en la creación de software usando recursos de *Design Thinking*: adaptación de recursos

Corso Cynthia, Challiol Cecilia, Paz Menvielle Alejandra, Ligorria Karina, Brito Carolina.

Centro de Investigación, Desarrollo y Transferencia de Sistemas de Información
Departamento Ingeniería en Sistemas de Información
Centro de Investigación y Transferencia Acústica
Facultad Regional Córdoba. Universidad Tecnológica Nacional
Maestro M. López esq. Cruz Roja-Ciudad Universitaria-Córdoba
cynthia@bbs.frc.utn.edu.ar/ceciliac@lifia.info.unlp.edu.ar/
mpaz@frc.utn.edu.ar/kligorria@frc.utn.edu.ar/carobrito@gmail.com

RESUMEN

El presente trabajo explora la adaptación necesaria de los recursos de *Design Thinking* (DT) dentro de un abordaje de *Diseño Participativo* (DP) para la creación de software; haciendo foco en cómo se debe adaptar la dinámica de estos recursos de una modalidad presencial a una virtual. El DP promueve la colaboración entre los participantes para desarrollar soluciones tecnológicas alineadas con las necesidades reales de los usuarios. Mediante el uso de DT, no solo se busca una solución funcional, sino también innovadora, que mejore la experiencia de los usuarios. La mayoría de los recursos DT surgieron para ser usados en una modalidad presencial, y actualmente requieren ser adaptados para ser efectivos en una modalidad virtual. El objetivo de este trabajo es comenzar a explorar cómo adaptar los recursos de DT para pasar de una modalidad presencial a una experiencia virtual.

Palabras claves: *Diseño Participativo, Design Thinking, Desarrollo de Software.*

CONTEXTO

Este trabajo se enmarca en el proyecto de investigación "Catálogo y recomendaciones de recursos (métodos) para facilitar experiencias de *Diseño Participativo* usando *Design Thinking*" (PID- SITCCO0008594), homologado por la Secretaría de Ciencia, Tecnología y Posgrado de la Universidad Tecnológica Nacional. El contexto de desarrollo de la presente investigación es el Centro de Investigación, Desarrollo y Transferencia de Sistemas de Información (CIDS) radicado en la U.T.N Facultad Regional Córdoba.

1. INTRODUCCIÓN

El desarrollo de software es un proceso que requiere una comprensión profunda de las necesidades de los usuarios finales. Sin embargo, en los últimos años ha cobrado mayor relevancia la idea de que los usuarios finales no deben ser vistos simplemente como receptores de la tecnología, sino como participantes activos en el proceso de creación. En este contexto, enfoques como el *Diseño Participativo* (DP) han ganado protagonismo, destacando la colaboración

entre diseñadores y usuarios para generar soluciones más efectivas y adaptadas a las necesidades reales.

Según Sanders (2002), el DP implica la participación de los usuarios finales en todas las etapas del proceso de diseño, desde la identificación de necesidades hasta la creación de soluciones [1]. Esto sugiere que la participación de los usuarios no debe limitarse a la recolección de datos o pruebas de usabilidad, sino que deben ser considerados co-creadores a lo largo de todo el proceso de diseño. A través de este enfoque, se busca empoderar a los usuarios, asegurando que sus voces sean escuchadas y que sus experiencias influyan directamente en la creación de soluciones tecnológicas.

Es fundamental que la solución de software resultante no solo sea funcional y se ajuste a las necesidades de los usuarios, sino que también sea innovadora. La innovación en el software no solo mejora la experiencia del usuario, sino que también introduce nuevas formas de interacción, aumenta la eficiencia y permite que la tecnología se adapte a nuevos contextos.

El DP no se limita a una única manera de ponerlo en práctica, sino que existen diversas formas de implementarlo. En este sentido, Maun et al. (2021) destacan que el enfoque de *Design Thinking* (DT) podría ser particularmente adecuado para llevar a cabo el DP, debido a su potencial para generar soluciones innovadoras [2]. Una de las definiciones más reconocidas de DT [3] es la propuesta por Cross (2011), quien lo describe como el conjunto de procesos cognitivos implicados en el trabajo de diseño, así como las habilidades y prácticas que los diseñadores emplean para crear nuevos artefactos o ideas y resolver problemas.
No obstante, es fundamental que los recursos [4][5][6] empleados dentro de este marco consideren las necesidades y demandas específicas de los usuarios finales. La selección adecuada de los recursos de *Design Thinking* en las experiencias de *Diseño Participativo* para la creación de software es crucial, ya que debe garantizar la participación activa de todos los involucrados.

En la etapa pre pandemia COVID19, DT era puesto en práctica en talleres mayoritariamente presenciales; sin embargo, la situación de pandemia abrió la posibilidad de llevarlos a cabo de manera virtual. Esto abrió también la incertidumbre de cómo adaptar los mismos para la modalidad virtual.

A partir de lo expuesto anteriormente, el objetivo de este trabajo es comenzar a explorar cómo adaptar los recursos de DT para pasar de una modalidad presencial a una experiencia virtual.

2. LINEAS DE INVESTIGACIÓN y DESARROLLO

En este proyecto de investigación convergen diversas líneas de investigación, por un lado, la *Ingeniería de Software*, y además dos áreas claves *Diseño Participativo* y *Design Thinking*. A continuación, se detallan los temas específicos que se abordan dentro de cada una de estas líneas de investigación.

- *Ingeniería de Software*
  - *Metodologías de Ingeniería de Software:* Esto permitirá poder considerar otros abordajes para poder tomar de ahí las mejores técnicas de diseño.

- Con relación a la temática de *Design Thinking se investigará:*
  - Conceptos básicos de *Design Thinking* (Empatía, Ideación, Prototipado, Evaluación)
  - Posibles frameworks para su puesta en práctica
  - Recursos posibles de ser usados en esta temática
  - Adaptación de recursos para ser usados con personas que tienen diferentes preferencias de participación
  - Rol del facilitador [7]
- Con relación a la temática de *Diseño Participativo* se investigará:
  - Conceptos básicos de Diseño Participativo
  - Abordajes posibles para llevarlo a cabo para reconocer las ventajas de cada uno.
  - Recursos usados en este contexto
  - Rol del facilitador [7]

3. RESULTADOS OBTENIDOS/ESPERADOS

Existen muchos y variados recursos que se utilizan en DT [8]. En la etapa pre pandemia COVID19, DT era puesto en práctica en talleres presenciales; sin embargo, la situación de pandemia abrió la posibilidad de llevarlos a cabo de manera virtual. Esto genero tener que empezar a adaptar su dinámica, aun así, todavía no hay un consenso de cómo hacerlo.

A continuación, se describe un recurso que tiene una dinámica presencial muy particular que requiere repensarlo para una modalidad virtual, este recurso es el *Brainwriting* [9] con una modalidad grupal.

El objetivo del *Brainwriting* es generar ideas de manera grupal y registrar las mismas de forma escrita. Una de las principales ventajas de este recurso es que asegura una participación equitativa de todos los integrantes del grupo. Además, las ideas generadas por unos participantes pueden motivar la aparición de nuevas ideas por parte de otros. Es un recurso que se usa generalmente en la etapa de idear de los frameworks de DT.

Antes de la experiencia presencial de DT, el facilitador tiene que decidir cuántas rondas se van a realizar y su duración, y debe preparar hojas para los participantes de cada grupo. Al momento de la experiencia de DT acontece lo siguiente:

- Los participantes de cada grupo sentados alrededor de una mesa reciben una hoja cada uno.
- El facilitador inicia la actividad. Cada participante de cada grupo escribe ideas en las hojas que reciben.
- Cuando el facilitador marca el fin de la primera ronda, cada hoja rota al compañero de la izquierda.
- Cuando cada participante recibe la nueva hoja, puede:
  - leer o no las ideas previamente generadas,
  - modificar ideas generadas anteriormente, y/o
  - escribir directamente sus ideas.
- El proceso de cada ronda es siempre igual, y es marcado por el facilitador. Esta actividad finaliza cuando el facilitador marca el tiempo de finalización de la última ronda.

A continuación, se analiza cómo un recurso pensado puramente para una modalidad presencial puede ser llevado a una modalidad virtual, para aprovechar su ventaja de participación equitativa.

Al momento de diseñar la actividad virtual, el facilitador tiene que decidir

cómo va a similar las hojas de manera virtual, por ejemplo, con una hoja distinta en un archivo digital el cual es compartido por grupos.

Al momento de la experiencia virtual, para simular lo mismo que acontece en el presencial todos deben de estar conectados sincrónicamente. Por otro lado, en la modalidad virtual, la hoja no se mueve como en el presencial, sino que son los participantes los que se mueven en las hojas del archivo digital. El mayor desafío es que los participantes entiendan esta dinámica, y no se mareen entre las hojas, ya que podrían estar escribiendo en una hoja que no les corresponde. Esto en una dinámica presencial no acontece porque cada persona recibe su hoja en un momento dado.

Más aún, la dinámica virtual del *Brainwriting* podría permitir trabajar de manera asincrónica, y cada uno estar trabajando ideas en una hoja virtual puntal, pero teniendo más tiempo para pensar ideas. Esta es una posibilidad que abre la modalidad virtual.

Otro recurso que se usa en las experiencias presenciales de DT es la matriz de 2x2 para clasificar ideas en función de dos dimensiones establecidas. Una de las principales ventajas es que ayuda a identificar visualmente una determinada relación de las ideas entre las dimensiones especificadas. Además, implica un acuerdo del grupo para determinar en qué cuadrante de la matriz se clasifica cada idea. Es un recurso que se usa generalmente en la etapa de idear de los frameworks de DT.

Para la matriz de 2x2, el facilitador en la etapa de diseño del recurso decide: la duración de la actividad, si es que plantea un tiempo máximo o no; cuáles serán los ejes de la matriz 2x2 (estos tienen un extremo positivo y negativo; y además generalmente da la libertad para que los grupos puedan auto-gestionar la forma de clasificar. Además, decide el tipo de material a usar para plasmar la matriz 2x2, por ejemplo, un afiche (en el cual dibuja la matriz).

Al momento de la experiencia de diseño, cada grupo recibe la matriz 2x2 (generalmente las ideas están previamente generadas están plasmadas en post-its). Cada grupo se auto-gestiona para determinar cómo clasificar las ideas, podría acontecer, por ejemplo:

- cada participante decide en que cuadrante ubicar sus ideas. Luego, el grupo acuerda si es la mejor ubicación o la reacomoda;
- cada grupo va analizando de una idea a la vez, y acuerda la mejor ubicación.

Cuando la matriz de 2x2 se quiere usar en una modalidad virtual, se debe representar la misma en un archivo digital, y requiere que los participantes estén conectados sincrónicamente. Sin embargo, el resto de las consideraciones son las misma que el presencial. En este caso, parece evidenciarse que solo cambia el medio donde se expresa la clasificación de ideas.

Estos dos recursos explorados muestran cómo se deben empezar a repensar en algunos casos las dinámicas de los recursos, pero además poder aprovechar las ventajas de la modalidad virtual que permite el trabajo tanto sincrónico como asincrónico en algunos casos, y potenciando el beneficio de cada uno. En el proyecto se seguirán explorando otros recursos de DT para analizar cómo deben ser adaptadas las dinámicas para ser llevados a una dinámica virtual.

4. FORMACIÓN DE RECURSOS HUMANOS

El proyecto contempla la formación de los recursos humanos involucrados, quienes, al participar en él, también contribuirán a su propio desarrollo profesional. En el contexto de esta investigación, se está llevando a cabo la tesis de Doctorado en Ciencias Informáticas de la directora del proyecto. Se integrarán al equipo docentes e investigadores de la UTN-Facultad Regional Córdoba, quienes se desempeñan en el rol de investigador de apoyo, con el objetivo de fortalecer su formación en investigación científica y tecnológica. Asimismo, participan becarios estudiantes de investigación, quienes colaboran en el desarrollo del marco metodológico.

Los estudiantes tendrán la oportunidad de realizar la Práctica Supervisada de quinto año en el marco de este proyecto. Los avances y propuestas generados estarán disponibles para su transferencia y aplicación en el Centro de Investigaciones, Desarrollo y Transferencia de Sistemas de Información (CIDS).

5. REFERENCIAS

[1] Sanders, E. B. N. (2002). From user-centered to participatory design approaches. Design and the Social Sciences, pp. 18-25. CRC Press.

[2] Maun, R., Fabri, M. Trevorrow, P. (2021). Adapting Participatory Design Activities for Autistic Adults: A Review. International Conference on Human-Computer Interaction (HCI), Niagara Falls, Canada, pp. 300-314. https://doi.org/10.1007/978-3-030-78224-5_21

[3] Cross, N. (2011). Design thinking: Understanding how designers think and work. Berg. Oxford. New York.

[4] Baella.com. 22 Herramientas de Design Thinking fase a fase. https://baella.com/noticias/22-herramientas-de-design-thinking-fase-a-fase (último acceso: 01/02/2025).

[5] Design Thinking en España. Herramientas de Design Thinking. https://xn--designthinkingespaa-d4b.com/herramientas-de-design-thinking (último acceso: 13/02/2025).

[6] Thinkernautas. 34 herramientas para el Design Thinking en un rápido vistazo. https://thinkernautas.com/34-herramientas-para-el-design-thinking-en-un-rapido-vistazo (último acceso: 11/02/2025).

[7] Mosely, G., Markauskaite, L. y Wrigley, C. (2021). Design facilitation: A critical review of conceptualisations and constructs. *Thinking skills and creativity*, *42*, 100962.

[8] Hehn, Jennifer; Uebernickel, Falk; Herterich, Matthias. (2018). Design Thinking Methods for Service Innovation - A Delphi Study-. Pacific Asia Conference on Information Systems. (PACIS 2018), Japón, Art. 126. https://aisel.aisnet.org/pacis2018/126/

[9] Heslin, P. A. (2009). Better than brainstorming? Potential contextual boundary conditions to brainwriting for idea generation in organizations. *Journal of Occupational and Organizational Psychology*, *82*(1), pp. 129-145.

Framework de desarrollo de aplicaciones distribuidas para dispositivos ubicuos basado en WS-CDL

Oscar Testa[1]; Rubén Pizarro[1]; Pablo García[1]; Leandro Castro[1]; Sonia Ponteprimo[1]

[1] Facultad de Ciencias Exactas y Naturales
Universidad Nacional de La Pampa
Av. Uruguay 151 – (6300) Santa Rosa – La Pampa – Argentina
Tel.: +54-2954-425166– Int. 28
[otesta, ruben,pablogarcia,castro.leandro]@exactas.unlpam.edu.ar

Resumen

Este proyecto de investigación, surge como continuidad del proyecto "Ingeniería de Software: Composición de Servicios en Ambientes Ubicuos", desarrollado también en el ámbito de la Facultad de Ciencias Exactas y Naturales de la UNLPam.

Tomando como resultado el framework de coordinación de dispositivos ubicuos, basado en la utilización de los estándares de SOA para servicios web, se plantea como líneas futuras de investigación implementar el citado framework en ambientes reales, donde se pueda poner a prueba dicho entorno.

Se espera implementar el framework en una Huerta Tecnológica Escolar, como una estrategia pedagógica que contribuye en muchos aspectos a mejorar la calidad de vida de los alumnos participantes, como así también la de la comunidad en la que se encuentra inserta la Escuela Secundaria que realizará la experiencia. Además de poder implementar en un ambiente real el framework para coordinar los distintos sensores y actuadores para automatizar las tareas de riego y control de temperatura y humedad del cultivo.

Se está trabajando en el diseño de una implementación real del framework de coordinación de dispositivos, el cual abre las puertas para una producción de este tipo de soluciones para todo tipo de ambientes.

Palabras clave: sistemas ubicuos, SOA, servicios, composición de servicios, coreografías.

Contexto

El presente trabajo se enmarca en el Proyecto de Investigación: Coordinación de dispositivos de IoT para resolver problemas específicos de la distribución de energía eléctrica – Facultad de Ciencias Exactas y Naturales, Universidad Nacional de La Pampa. Las líneas aquí presentadas actualmente forman parte de las bases de un anteproyecto de tesis doctoral en ingeniería de software por la Universidad Nacional de San Luis, además de un proyecto de Tesis de Maestría en Enseñanza en Escenarios Digitales en AUSA.

Introducción

Los avances de las comunicaciones entre dispositivos ha permitido que estos sean generadores y consumidores de servicios al mismo tiempo, es decir, de acuerdo a las capacidades del dispositivo puede no solo obtener, sino también ofrecer a otros equipos sus funciones y así cooperar entre ellos. La tendencia actual es hacia los ambientes ubicuos, los cuales se caracterizan por estar poblados de numerosos dispositivos que, gracias a la integración extrema de los elementos electrónicos, son invisibles al usuario y están en permanente rastreo de la actividad humana [2].

Dispositivos ubicuos son todos aquellos dispositivos que pueden existir en todas partes, es decir, son dispositivos electrónicos que tienen capacidad de procesamiento y

comunicación y pueden ser encontrados en lugares diversos de la vida cotidiana.

La computación ubicua es un desarrollo tecnológico que intenta que las computadoras no se perciban en el entorno como objetos diferenciados, y que la utilización por parte de los seres humanos sea lo más transparente y cómoda posible, facilitando de esta manera la integración en la vida cotidiana. Desde hace varios años los dispositivos ubicuos han ganado importancia y presencia en la vida cotidiana de las personas, debido principalmente a que: poseen distintos tipos de sensores (posicionamiento, proximidad, luminosidad, temperatura, etc.), facilitan la conectividad incluso en áreas con poca señal o acceso a las redes, permiten la convergencia tecnológica (computo, medios, telefonía, etc) y brindan acceso a servicios de distinta índole (mapas, ayudas, etc).

Por composición entendemos la forma en que se pueden combinar o enlazar un número indeterminado de dispositivos para llevar adelante una tarea determinada. En ambientes ubicuos, la composición de dispositivos, presenta nuevos desafíos tales como: la heterogeneidad (ya sea por la diversidad de dispositivos involucrados, como por la presencia de dispositivos de varios fabricantes), las contingencias de los dispositivos y la personalización de los mismos (por ej. provisión de servicios de acuerdo a las preferencias del usuario). Dado que los dispositivos en donde los servicios son ejecutados poseen limitaciones de recursos (ej. poca memoria y batería), se deben hacer consideraciones especiales respecto a la eficiencia y rendimiento de la composición de servicios [3].

La composición en este tipo de ambientes implica que los dispositivos deben dialogar entre ellos para poder compartir los servicios que ofrecen con la finalidad de obtener un servicio con valor agregado, o bien para abordar la solución de una problemática particular, como podría ser la seguridad de un hogar, o la seguridad vial, por mencionar algunos ejemplos.

Si bien hoy en día podemos decir que distintos sensores o dispositivos se pueden comunicar entre ellos, compartiendo de alguna manera sus servicios, generalmente lo realizan a partir de protocolos propietarios y sin seguir definiciones estándares, provocando que otros componentes de otros proveedores (o incluso de los mismos) no puedan ser utilizados. Esto obviamente representa una importante limitación en la composición de dispositivos ubicuos. Adicionalmente la composición de dispositivos ubicuos presenta un nuevo desafío. Los mecanismos de composición en ambientes masivos, necesitan hacer frente a las distintas contingencias que pueden ocurrir con estos dispositivos. Los dispositivos ubicuos tienen distintas limitantes como son la cantidad de memoria disponible, la durabilidad de la batería, la disponibilidad de acuerdo a la red del lugar donde se encuentre en un momento determinado. Todas estas variantes hacen que la composición de dispositivos[1] ubicuos se transforme en un área de investigación muy importante donde los avances no han sido claros al día de hoy[3].

Líneas de Investigación y Desarrollo

La orientación a servicios proporciona estrategias bien conocidas (por ejemplo, las orquestaciones) para la composición de servicios. Tales estrategias resultan herramientas de gran utilidad para la construcción de aplicaciones complejas mediante la combinación de servicios heterogéneos, los cuales presentan la característica muy favorable de su simplicidad. Además, su carácter distribuido los convierte en una solución especialmente aplicable a

[1] Si bien los autores se refieren a la composición de servicios, se hace dentro de un contexto de dispositivos ubicuos, lo cual a los fines de este trabajo se puede interpretar como composición de dispositivos, haciendo que la terminología para este caso particular sea más adecuada.

ambientes ubicuos. Sin embargo, en contextos donde los servicios son dinámicos, móviles, menos fiables y dependientes del dispositivo, los mecanismos de composición establecidos para servicios web no pueden aplicarse directamente [8].

Por otro lado, la composición de múltiples dispositivos ubicuos presenta características propias que la separan de la composición tradicional de servicios web. Es necesario lidiar con las características heterogéneas que cada uno de ellos podría presentar, como por ejemplo la cantidad de memoria y la duración de la batería.

Los dispositivos ubicuos se conectan en contextos diferentes en cada ocasión y, en consecuencia, su disponibilidad es muy variable y la confiabilidad se vuelve extremadamente dinámica. Todas estas dificultades hacen que la composición de dispositivos se transforme en un área de investigación que aún presenta muchos aspectos a dilucidar y que resulta de interés en términos de investigación [3].

En la actualidad, existen múltiples proyectos donde se integran sensores y dispositivos ubicuos a situaciones problemáticas del mundo real. Por ejemplo, la domótica busca que dispositivos y sensores deban actuar en coordinación para prevenir incidentes de seguridad diversos en contextos domésticos.

El concepto de Industria 4.0 [9], busca integrar dentro de una planta fabril la intercomunicación de todos los dispositivos que componen la cadena de producción, buscando, fundamentalmente, una coordinación de tareas que atienda las variables que afectan al proceso productivo (tiempos, stocks, demandas).

Es inmediato detectar múltiples tipos de procesos en los cuales los dispositivos ubicuos adquieren trascendencia: aeronáutica, automotriz, alimentos, etc.

En consecuencia, esta investigación en particular aborda la implementación de un framework de coreografía de servicios con dispositivos ubicuos, la cual está definida por Testa, et all. en[11]. A partir de la utilización de este framework es que se propone implementarlo sobre sistemas reales, y específicamente para automatizar las tareas de riego, temperatura y humedad ambiente y del suelo en una Huerta Tecnológica Educativa, en un ambiente educativo de la ciudad de Dorila, Provincia de La Pampa.

Resultados y Objetivos

La proliferación de dispositivos de computación ubicuos e interconectados (PDAs, tabletas, móviles, etc), así como los recientes avances en la tecnología de radio frecuencia y las redes de sensores están fomentando la creación de ambientes donde las aplicaciones de internet y los servicios se están haciendo más populares y necesarias para los usuarios de móviles. La composición de servicios a través de múltiples dispositivos móviles presenta un nuevo desafío el cual no es compatible con la composición de servicios como se plantea actualmente. En particular, los mecanismos de composición en ambientes masivos, como lo es el de dispositivos móviles, necesita hacer frente a las distintas contingencias que pueden ocurrir con estos elementos (desconexión, falta de batería, etc), así como también contemplar la heterogeneidad de los mismos. Estos dispositivos tienen distintas limitantes como son la cantidad de memoria disponible, la durabilidad de la batería, la disponibilidad de acuerdo a la red del lugar donde se encuentre en un momento determinado. Todas estas variantes hacen que la composición de servicios incluyendo dispositivos móviles se transforme en un área de investigación muy importante donde los avances no han sido claros al día de hoy [3].

En este tipo de ambientes, mecanismos automáticos y dinámicos son necesarios para la composición de servicios, ya que de esta forma se puede compensar la falta de disponibilidad de un servicio en un momento determinado [8].

Si bien a este problema existente se encontró solución, a través de la implementación de un framework de ejecución de coreografías en ambientes ubicuos, tal como se plantea en [11],

se necesita hacer pruebas y más investigaciones a partir de una implementación real.

Por lo antes expuesto se deducen las siguientes preguntas de investigación:

- ¿El framework de ejecución de coreografías propuesto, brinda soluciones en un ambiente real de trabajo?
- ¿Se necesitan mejoras para ser implementado en un ambiente real? ¿Cuáles?
- ¿Qué tipo de hardware se necesita para una implementación real, o alcanza con utilizar placas académicas de uso general como Arduino?
- ¿Los protocolos de comunicación seleccionados, son los correctos, o debe ampliarse la implementación y utilización de protocolos?

Por lo tanto, el proyecto plantea los siguientes objetivos:

- Realizar una implementación real del framework de ejecución de coreografías en ambientes ubicuos, en un caso de estudio particular y real.
- Realizar las mejoras y ajustes necesarios al framework para que pueda ser utilizado comercialmente.
- Poner a prueba la validez de la utilización de los protocolos de comunicación establecidos por SOA.
- Difundir los resultados obtenidos con publicaciones de los resultados parciales y finales.

La hipótesis principal es poner a prueba si los mecanismos principales y fundacionales de SOA son aplicables para la composición de coreografías en ambientes ubicuos, en un ambiente real. Se tienen fuertes indicios, por los estudios e investigaciones llevadas adelante en el proyecto de Investigación sobre el cual se basa el presente trabajo, que los mismos serán suficientes y que serán de utilidad en este tipo de ambientes.

Se espera, además logros en términos de formación en recursos humanos:

El Mg. Pablo M. García, trabaja en la propuesta de innovaciones en el ámbito del manejo de la memoria de las placas arduino, cuyos resultados deben reflejarse en una Tesis Doctoral.

La investigadora Sonia Ponteprimo se encuentra trabajando en la implementación del framework en la Huerta Tecnológica, donde se están haciendo los primeros análisis de cantidad de dispositivos a ubicar, lugares y qué actuadores se utilizarán para las tareas de automatización del riego.

Como parte de los resultados obtenidos hasta el momento podemos mencionar que se ha obtenido una precisa y clara situación actual o estado del arte respecto de los dispositivos ubicuos que formarán parte de la implementación real, así como los distintos avances que se han realizado en la teoría de microservicios y de coordinación a través de la utilización de coreografías de servicios. Además se ha logrado la implementación en el framework, de un nuevo protocolo de comunicación de los dispositivos como el caso de Bluetooth. También se ha logrado armar un prototipo real para la implementación en la Huerta Tecnológica, el cual sirve como modelo para su implementación real.

Formación de Recursos Humanos

Además de las metas propuestas en el punto 3, se espera como resultado en la formación de recursos humanos, la continuación de esta misma línea de proyecto como tesis doctoral del investigador Mg. Pablo García y una tesis de Maestría para la investigadora Sonia Ponteprimo, relacionada con las metodologías de enseñanza en espacios digitales sobre la aplicación en la Huerta Tecnológica Educativa. También se espera lograr una mayor interrelación con la Universidad Politécnica de Madrid para la aplicación de las metodologías aquí presentadas en los proyectos de Ingeniería de Software Empírica y realizando una estancia posdoctoral del Dr. Oscar A. Testa. Adicionalmente, se espera que otras tesis de

Maestría surjan a partir de los logros obtenidos en la presente línea de investigación.

Bibliografía

[1] U. I. d. T. (UIT), "Unión Internacional de Telecomunicaciones." https://www.itu.int/net/pressoffice/pressreleases/2014/23-es.aspx, 10 2015.

[2] M. Weiser, "Hot topics-ubiquitous computing," Computer, vol. 26, pp. 71–72, Oct 1993.

[3] Q. Z. Sheng, X. Qiao, A. V. Vasilakos, C. Szabo, S. Bourne, and X. Xu, "Web services composition: A decade's overview," Information Sciences, vol. 280, no. 0, pp. 218–238, 2014.

[4] M. Viroli, "On competitive self-composition in pervasive services," Science of Computer Programming, vol. 78, no. 5, pp. 556–568, 2013. Special section: Principles and Practice of Programming in Java 2009/2010 & Special section: Self-Organizing Coordination.

[5] S. W. Loke, "Supporting ubiquitous sensor-cloudlets and context-cloudlets: Programming compositions of context-aware systems for mobile users," Future Generation Computer Systems, vol. 28, no. 4, pp. 619–632, 2012.

[6] F. Palmieri, "Scalable service discovery in ubiquitous and pervasive computing architectures: A percolation-driven approach," Future Generation Computer Systems, vol. 29, no. 3, pp. 693–703, 2013. Special Section: Recent Developments in High Performance Computing and Security.

[7] S. Najar, M. K. Pinheiro, and C. Souveyet, "A New Approach for Service Discovery and Prediction on Pervasive Information System," Procedia Computer Science, vol. 32, pp. 421–428, 2014. The 5th International Conference on Ambient Systems, Networks and Technologies (ANT-2014), the 4th International Conference on Sustainable Energy Information Technology (SEIT-2014).

[8] G. Cassar, P. Barnaghi, W. Wang, S. De, and K. Moessner, "Composition of services in pervasive environments: A Divide and Conquer approach," in Computers and Communications (ISCC), 2013 IEEE Symposium on, pp. 000226–000232, July 2013.

[9] Wikipedia, "Industria 4.0 — Wikipedia, La enciclopedia libre," 2016. [Internet; descargado noviembre-2016].

[10] H.-I. Yang, R. Bose, A. (Sumi) Helal, J. Xia, and C. Chang, "Fault-Resilient Pervasive Service Composition," in Advanced Intelligent Environments (A. D. Kameas, V. Callagan, H. Hagras, M. Weber, and W. Minker, eds.), pp. 195–223, Springer US, 2009.

[11] Oscar A. Testa, Efraín R. Fonseca C., Germán Montejano and Oscar Dieste, "Coordination of Ubiquitous Devices in Pervasive Environments: A Proposal Based on WS-CDL" in 38th International Conference of the Chilean Computer Science Society. SCCC 2019.

Hacia la Gestión de Activos en Sistemas de Big Data basado en CBR orientado a Servicios

Andres Flores, Agustina Buccella, Alejandra Cechich, Alan De Renzis
GIISCo Research Group
Departamento de Ingeniería de Sistemas – Facultad de Informática
Universidad Nacional del Comahue
Neuquén, Argentina
Contacto: [andres.flores, agustina.buccella]@fi.uncoma.edu.ar

RESUMEN

El desarrollo de Sistema de Big Data (SBD) involucra grandes desafíos. La diversidad de los datos involucra distintas características, o Variedad: estructural, de las fuentes, de contenido y de procesamiento. Identificar estas variedades genera desafíos de modelado que permitan una amplia reusabilidad: tanto de los datos como en la gestión del proceso de reuso. El proyecto de investigación actual propone un Proceso de Modelado de Variedad y Reuso de Activos en SBD, para identificar aspectos relevantes de variedad en distintos casos de analítica de datos. Se está desarrollando una plataforma que soporte el proceso de identificación, almacenamiento y reuso de casos de analítica de datos, incorporando capacidades de recomendación, trazabilidad y visualización. En un proyecto de investigación previo, enfocado en Computación Orientada a Servicios, se definió un Proceso de Razonamiento basado en Casos (CBR) para Descubrimiento y Selección de Servicios Web. La combinación de ambas investigaciones implicaría considerar que los activos de dominio de SBD podrían ser encapsulados como servicios, para ser tratados como casos en el proceso de CBR, y posteriormente ante la necesidad de analizar un caso nuevo, realizar descubrimiento de casos, y analizar los activos de dominio almacenados para encontrar oportunidades de reuso, tanto de activos de procesamiento como de conocimiento (y analítica de datos).

***Palabras Clave*:** Big Data – Variedad Reuso – Software Orientado a Servicios.

CONTEXTO

La línea presentada se inserta en el contexto del Proyecto 04/F019 "Modelado de Variedad en Sistemas Big Data". Financiado por UNCo. (2022-2026).

INTRODUCCIÓN

Los Sistemas Big Data (SBD) son sistemas inherentemente distribuidos, cuyas tecnologías y frameworks evolucionan rápidamente, lo cual genera grandes desafíos para arquitectos y diseñadores, particularmente sobre los requerimientos que una arquitectura para SBD debe cumplir; entre ellos, considerar las cinco "Vs" (*Volumen*, *Velocidad*, *Variedad*, *Variabilidad* y *Veracidad*) [DL20]. Un cambio importante con respecto a depósitos de datos tradicionales, es la naturaleza no estructurada de algunos datos que pueden provenir de diversas fuentes: sensores, redes sociales, entorno y la propia empresa cliente. La diversidad de esos datos puede analizarse considerando distintas características, lo cual en SBD se denomina *Variedad*, y es posible categorizarlas de acuerdo a una taxonomía que divide el análisis de variedad en cuatro casos de diversidad: *estructural*, de las fuentes, de *contenido* y de *procesamiento* [AJ15]. La diversidad *estructural* denota la variedad en formatos y tipos de datos, clasificándolos como estructurados, semi-estructurados y no

estructurados; la diversidad de las *fuentes* se clasifica en tres grupos – datos generados por humanos (resultados de interacciones hombre-computador), generados por máquinas (automáticamente a partir de dispositivos, como sensores), o mediados por procesos (correlación con datos generados por humanos); la diversidad de *contenido* considera diferentes tipos de soporte (único medio, multimedio o gráfico); y la diversidad de *procesamiento* se enfoca en las distintas necesidades de procesamiento algorítmico (batch, interactivo, streaming o gráfico). Identificar la variedad en SBD abre el desafío a alternativas de modelado, sobre todo analizando posibilidades de reusabilidad futura de los aspectos comunes (no variantes) de un SBD. En ese sentido, han surgido diferentes propuestas que enfocan en aspectos diversos del reuso – tanto de los datos como de la gestión del proceso de reuso. Un reuso efectivo requiere de un soporte que permita almacenar, identificar y eventualmente reusar recursos.

En este proyecto estamos desarrollando un Proceso de Modelado de Variedad y Reuso de Activos en SBD [BCMetal24], para lo cual actualmente se efectúan aplicaciones en el dominio de la agricultura inteligente, en colaboración con INTA Alto Valle, Río Negro. El proceso alcanzado permite identificar aspectos relevantes de variedad en distintos casos de analítica de datos, y se espera almacenar diferentes activos que faciliten el reuso en problemas similares. Con lo cual se está desarrollando una plataforma que soporte el proceso de identificación, almacenamiento y reuso de casos de analítica de datos, incorporando capacidades de recomendación, trazabilidad y visualización.

En un proyecto previo [FGCetal21] se enfocó la investigación en el paradigma de Computación Orientada a Servicios (SOC), que promueve fuertemente el reuso de software. La funcionalidad a ser reusada adopta la forma de servicios, o unidades lógicas que presentan entornos heterogéneos de ejecución y pueden ser ensambladas para formar otras unidades lógicas de mayor nivel de abstracción que resuelvan (directamente o en parte) los procesos de negocios para un contexto de aplicación [PTDL07]. El paradigma SOC encontró una plataforma potencial de aprovechamiento mediante la Web, desde donde se desarrolló la tecnología de Servicios Web [Wetal05], con notaciones formales específicas para la descripción las interfaces de servicios. Los Servicios Web pueden adoptar diferentes estilos: SOAP y RESTful [GFMCZ16]. Los servicios SOAP se describen mediante WSDL (Web Service Description Language) para especificar las operaciones e intercambio de mensajes. Existen diferentes versiones del lenguaje WSDL (1.0, 1.1, 2.0), siendo la versión WSDL 2.0 un estándar OMG. Los servicios RESTful se pueden describir mediante diferentes lenguajes, siendo WADL (Web Application Description Language) el primer intento de notación como analogía de WSDL. La industria ha propuesto otros lenguajes: OpenAPI/ Swagger, REST API Modeling Language (RAML), o MIT's API Blueprint.

Evaluar y consumir tal variedad de servicios heterogéneos podría reducir el espacio de solución de los servicios a consumir, de acuerdo con la tecnología que se pueda soportar. De este modo, la amplia gama de servicios probablemente reusables, se vuelve seriamente limitada.

Para ello, hemos desarrollado un Metamodelo de Servicios Heterogéneos basado en estándares [GF19] para ampliar la gama de compatibilidad de servicios de diversas tecnologías, que pudieran ser potenciales candidatos para aplicaciones consumidoras. Los principales estándares para la descripción del servicio incluyen: WSDL 2.0, WADL y OpenAPI, considerando así ambos estilos de servicios, basados en SOAP y REST.

En función del Metamodelo de Servicios se ha construido un Proceso de

Razonamiento basado en casos (CBR) para el almacenamiento, descubrimiento y selección de servicios [FGCetal21]. CBR resuelve problemas utilizando o adaptando soluciones a problemas antiguos [RS13], gracias a la función de similitud utilizada para cuantificar la semejanza entre un par de casos, basados en servicios, evaluando las descripciones de la interfaz WSDL. El Metamodelo de Servicios permite representar servicios candidatos heterogéneos como casos reutilizables. Los nuevos casos se pueden resolver a partir de varios casos similares anteriores, que se recuperan mediante funciones de similitud basadas en diferentes algoritmos *K-nn* configurables. El proceso CBR junto con el Metamodelo de Servicios proveen soporte a la selección de servicios heterogéneos, agnósticos a la tecnología de una manera transparente. Esto aprovecha nuestro proceso CBR como un enfoque de gestión del conocimiento, para capturar y enriquecer el conocimiento obtenido de las selecciones sucesivas de servicios.

LÍNEAS DE INVESTIGACIÓN, Y DESARROLLO

Actualmente, abordamos aspectos sobre la efectiva detección y reuso de casos previos en analítica de datos, lo cual depende en gran medida de la representación de la documentación de los activos de dominio, incluyendo la variedad. El proceso de identificación desarrollado, es soportado actualmente por un primer prototipo, denominado COVAMAT, que requiere ser extendido para cumplir con los propósitos planteados [OCBetal23]. Entre esas extensiones se incluyen:

Reuso de casos: la versión actual de COVAMAT incluye una función de similitud para encontrar casos similares, de acuerdo a una consulta. Aunque la función está simplificada a una igualdad del contexto consultado y de los casos almacenados, que implica bajas capacidades de reuso de casos. Esta función debe ser ampliamente extendida en términos de estudios más avanzados sobre similitud, para identificar las variables más relevantes dentro de un dominio, que permiten determinar la similitud de casos; o para reconocer combinación de valores de contexto, que bajo ciertas condiciones, podrían presentar una similitud no inicialmente evidente.

Almacenamiento de recursos de los casos: hasta el momento COVAMAT posee una facilidad de almacenamiento acotada a los metadatos de los casos de dominio, considerando las variedades más relevantes. Por lo tanto, se requiere una gran extensión sobre el almacenamiento de diferentes activos del dominio con artefactos que permiten la definición de variedades para luego ser instanciadas al implementar un caso de dominio específico. De esta forma, para cada caso de dominio, la extensión requerida incluye agregar los mecanismos necesarios para almacenar: (1) los recursos que permitan identificar y recuperar las fuentes originales utilizadas, (2) las fuentes transformadas, es decir el resultado de las transformaciones durante el preprocesamiento, (3) los notebooks (códigos) trabajados, y (4) los resultados de las analíticas aplicadas. De esta forma, no sólo se almacenarían las variedades y las instanciaciones realizadas para casos de dominio, sino también los artefactos de software y recursos creados para su completo funcionamiento.

RESULTADOS OBTENIDOS / ESPERADOS

En [BCMetal24] hemos enumerado una serie de contribuciones previas, enfocadas en la identificación y modelado de variedad en SBD. El proceso desarrollado plantea dos enfoques para identificar variedad en un dominio: (1) identificación desde los datos o bottom-up (B-VIP) y (2) identificación desde los requerimientos o top-down (T-VIP). B-VIP considera la definición de un problema de un dominio e intenta detectar características variantes a partir de los datos existentes en uno o

varios repositorios. T-VIP considera el análisis de requerimientos y del dominio para contrastar suposiciones con los datos. Las variedades de dominio son documentadas considerando sus metadatos mediante una hoja de datos (datasheet), que permiten la identificación posterior para un posible reuso de casos, detectando variedades de contexto.

Para analizar el impacto que diversas características de contexto podrían tener en las inferencias que se realizan a partir de los datos, se han efectuado aplicaciones en colaboración con INTA Alto Valle [CBetal23, BCMetal24]: Identificación de variedad bottom-up para el análisis de calidad de cuerpos de agua superficiales; Caracterización de variables que influencian la turbidez de canales de riego; Caracterización de Variables para el Análisis del Índice de Vegetación; Caracterización de variables meteorológicas que influencian las fluctuaciones del nivel freático; entre otros.

Actualmente, la facilidad en la recolección y almacenamiento de datos, y las tecnologías e infraestructuras para manipularlos, ha motivado la creación de sistemas intensivos en datos (Big Data). En la práctica, el crecimiento de SBD obliga a plantear un cambio de paradigma de desarrollo: de dirigidos por la funcionalidad, a dirigidos por los datos. Es decir, el valor de los sistemas se asocia especialmente a la correcta identificación y explotación de los datos (y a su correspondiente calidad) [GK15].

En base a la posibilidad de reuso que provee el paradigma SOC, se han propuesto soluciones arquitectónicas para SBD con orientación a servicios, dado que la arquitectura de referencia estándar para SBD [NIST16] propone de manera conceptual una división de componentes funcionales lógicos conectados mediante interfaces, en la forma de servicios, que permiten interoperabilidad, y soportan la gran variedad en entornos de negocios, de una manera desacoplada, facilitando comprender que Big Data complementa y difiere de los enfoques existentes en analítica, inteligencia de negocios, bases de datos y sistemas. Se han propuesto algunas arquitecturas para SBD [GK15, KB19]: Lambda, Kappa, basada en Microservicios, y Zetta, entre otras. La arquitectura Lambda, se convirtió en un estándar en SBD, y está dividida en tres capas: la capa de procesamiento batch de datos históricos; la capa de interacción con el usuario; y la capa de velocidad para procesar flujos de datos en tiempo real. La arquitectura Kappa, elimina la capa batch, con capacidades avanzadas en la capa de velocidad. La arquitectura de Microservicios [WT24], ofrece la ventaja de computación en la nube, con una estructura desacoplada de servicios REST. La arquitectura Zetta, es una variante orientada a Microservicios, pero a un nivel más alto de granularidad, para interconectar arquitecturas de negocio o empresariales.

En el contexto del Proceso de Modelado de Variedad y Reuso de Activos en SBD, y su herramienta de soporte COVAMAT, el almacenamiento de recursos y activos de dominio podrían ser tratados mediante servicios en la nube, como versiones de las arquitecturas Lambda y de Microservicios. Tanto los notebooks o códigos para procesamiento de colecciones de datos, como las mismas colecciones de datos, con sus versiones (original, post-procesado, etc.) podrían ser encapsulados en la forma de servicios, para ser almacenados en la nube. En este sentido, el Proceso de CBR [FGCetal21] para razonamiento y selección de servicios heterogéneos, podría ser ajustado de forma pertinente al contexto de SBD, para actuar como un motor de descubrimiento de servicios, con la capacidad de almacenamiento de activos y recursos de los casos de dominio generados, en el proceso B-VIP; y luego con la potencialidad de la recuperación y reuso, para el proceso T-VIP. Por otro lado, se deberán abordar los desafíos relacionados con el reuso de conocimiento de dominio, que

plantea el proceso T-VIP en su totalidad. lo cual plantea nuevas oportunidades de investigación y desarrollo en SBD. La visión de esta línea de investigación se resume en:

"Extender el prototipo COVAMAT como plataforma orientada a servicios basada en CBR, que soporte el proceso de identificación, almacenamiento y reuso de casos de analítica de datos, y gestión de activos de software, con capacidades de recomendación, trazabilidad y visualización".

FORMACIÓN DE RECURSOS HUMANOS

El proyecto completo reúne aproximadamente a unos 15 investigadores, entre los que se cuentan docentes y estudiantes de UNComa, y colaboradoras expertas del dominio de aplicación, pertenecientes al INTA. Se cuenta con cuatro doctores y un magister. Tres de los docentes-investigadores están realizando carreras de postgrado en la temática del proyecto de investigación.

BIBLIOGRAFÍA

[AJ15] Abawajy, J. (2015). *Comprehensive analysis of big data variety landscape*. International Journal of Parallel Emerging Distributed Systems, 30(1):5–14.

[BCMetal24] Buccella, A., Cechich, A., Mazalu, R., Luzuriaga, J., Villegas, C., Cruz, C., Corgatelli, F., Martinez, R., Saurin, F., Torres, F. (2024). *Hacia una Plataforma para la Gestión de Variedad en Sistemas Big Data*. WICC, Workshop de Investigadores en Ciencias de la Computación, Santa Cruz.

[CBetal23] Cechich, A., Buccella, A., Villegas, C., Montenegro, A., Muñoz, A., and Rodriguez, A. (2023). *A model of reusable assets in AIE software systems*. Journal of Computer Science & Technology, 23(2):e13.

[DL20] Davoudian, D., Liu, M. (2020). *Big Data Systems: A Software Engineering Perspective*. ACM Computing Surveys, 53(5).

[FGCetal21] Flores, A., Garriga, M., Cechich, A., Moyano, M., De Renzis, A., Anabalon, D., Corgatelli, F. (2021). *Reuso de Servicios Heterogéneos basado en CBR*. WICC, Workshop de Investigadores en Ciencias de la Computación. La Rioja.

[GF19] Garriga, M., Flores, A. (2019). *Standards-driven Metamodel to Increase Retrievability of Heterogeneous Services*. SAC, ACM/SIGAPP Symposium on Applied Computing, pp. 2507-2514.

[GFMCZ16] Garriga, M., Flores, A., Mateos, C., Cechich, A., Zunino, A. (2016). *RESTful service composition at a glance: A survey*. Journal of Network and Computer Applications, 60:32-53.

[GK15] Gorton, I., Klein, J. (2015). *Distribution, data, deployment: Software architecture convergence in Big Data systems*. IEEE Software, 32(3):78–85.

[KB19] Kalipe, G., Behera, R. (2019). *Big Aata architectures: A detailed and application oriented review*. International Journal of Innovation, Technology and Exploratory Engineering, 8:2182–2190.

[NIST16] NIST (2016). NIST Big Data Interoperability Framework: vol6 Reference Architecture. Technical report, National Institute of Standards and Technology.

[OCBetal23] Osycka, L., Cechich, A., Buccella, A., Montenegro, A., and Muñoz, A. (2023). *COVAMAT: Functionality for variety reuse through a supporting tool*. Cloud Computing, Big Data & Emerging Topics, pp 57–74, Cham. Springer.

[PTDL07] Papazoglou, M.; Traverso, P.; Dustdar, S.; Leymann, F. (2007). *Service-Oriented Computing: State of the Art and Research Challenges*. IEEE Computer, 40(11): 38-45.

[RS13] Riesbeck, C., Schank, R. (2013). *Inside Case-based Reasoning*. Psychology Press.

[Wetal05] Weerawarana, S.; Curbera, F.; Leymann, F.; Storey, T.; Ferguson, D. (2005). *Web Services Platform Architecture: SOAP, WSDL, WS-Policy, WS-Addressing, WS-BPEL, WS-Reliable Messaging, and More*. Prentice Hall PTR.

[WT24] Werner, S., Tai, S. (2024). *A reference architecture for serverless Big Data processing*. Future Generation Computer Systems, 155:179–192.

INDUSTRIA 4.0: ESTUDIO DE HERRAMIENTAS SOFTWARE PARA LA EVALUACIÓN DE LA MADUREZ DIGITAL DE PYMEs

Nélida R. Cáceres, Ana C. Tolaba, José R. Quispe, Liliana M. Medina

Facultad de Ingeniería - Universidad Nacional de Jujuy
Ítalo Palanca 10 San Salvador de Jujuy – 0388 4221576
nrcaceres@fi.unju.edu.ar

RESUMEN

La industria 4.0, más conocida como la cuarta revolución industrial, se caracteriza por la informatización y digitalización de los procesos productivos y la generación de grandes cantidades de datos. Esta industria combina el hardware, software, tecnologías, personal humano capacitado en competencias digitales y políticas públicas adecuadas, con el fin de lograr la eficiencia de los procesos y mejorar la competitividad y sostenibilidad de las organizaciones en un contexto tecnológico. La madurez digital es fundamental para que las organizaciones puedan integrarse exitosamente en la Industria 4.0. La evaluación de la madurez digital permite identificar fortalezas y debilidades tecnológicas en todo tipo de organizaciones, a partir de las cuales es posible diseñar planes de acción específicos para su transformación digital.

Considerando este contexto, en este proyecto nuestro interés está centrado en estudiar diferentes herramientas de software utilizadas por empresas industriales, manufactureras y PyMEs que permiten realizar la medición y diagnóstico de su madurez digital. En particular, utilizar estas herramientas en PyMEs que se desempeñan en la provincia de Jujuy para diagnosticar su capacidad tecnológica. A partir de la evaluación realizada, será posible generar recomendaciones que permitan a las mismas implementar estrategias adecuadas para su transformación digital.

Palabras clave: *Industria 4.0, Madurez Digital, PyMEs.*

CONTEXTO

La propuesta de este trabajo, se encuentra enmarcada en el proyecto *"Estudio de Herramientas Software para la evaluación de la Madurez Digital: PYMEs de Jujuy camino hacia la Industria 4.0*", que fue presentado dentro de la convocatoria de proyectos de investigación Fortalecer - Integrar - Promover (FIP-UNJu 2025) y aprobado por la Secretaría de Ciencia y Técnica de la Universidad Nacional de Jujuy en la modalidad Promover (P), el cual se encuentra bajo incentivo.

1. INTRODUCCIÓN

En las últimas décadas, la Industria 4.0 ha emergido como una revolución en el ámbito industrial. El término "industria 4.0" se refiere a un nuevo modelo de organización y de control de la cadena de valor a lo largo del ciclo de vida del producto y de los sistemas de fabricación, apoyado y posibilitado por las tecnologías de la información. También es habitual referirse a este concepto con términos como "Fábrica Inteligente" o "Internet industrial". Todos estos términos tienen en común el reconocimiento de que los procesos de fabricación se encuentran en un proceso de transformación digital, una "revolución industrial" impulsada por el avance de las tecnologías de la información y, en particular, de la informática y el software [1].

El concepto de revolución industrial está ligado a cambios en las condiciones tecnológicas de producción, se lo asocia con la informatización y digitalización, y con la generación, integración y análisis de una gran cantidad de datos a lo largo del proceso

productivo y del ciclo de vida de los productos, facilitados fundamentalmente por las tecnologías de internet [2].

Entre los pilares tecnológicos de la industria 4.0 se destacan [3]: Internet de las cosas (IoT), Big data y análisis de grandes datos, Computación en la nube, Inteligencia Artificial, Ciberseguridad y Realidad Aumentada. Estas tecnologías se completan con otras como: sistemas de integración, máquinas y sistemas autónomos, manufactura aditiva, simulación de entornos virtuales, sensores inteligentes, controladores, plataformas electrónicas abiertas, sistemas de localización, sistemas de autoidentificación y blockchain.

El Plan de Desarrollo Productivo Argentina 4.0 [4] menciona que la mayor parte de las empresas todavía se encuentran operando con tecnologías anticuadas [5]. Muchas no tienen en claro los beneficios que pueden percibir por adoptar tecnologías 4.0, por lo que todavía se encuentran reacias a asumir el tiempo y los costos de la adaptación [6].

Nuestro país forma parte del conjunto de países en vías de desarrollo para los cuales la Revolución 4.0 representa un desafío de magnitud. La ausencia de planificación y políticas públicas activas para la adopción y desarrollo de soluciones 4.0 traería aparejados enormes riesgos en términos de atraso productivo, primarización económica y desempleo.

Es por ello que desde el Ministerio de Desarrollo Productivo de la Nación se decidió llevar adelante el Plan de Desarrollo Productivo Argentina 4.0, con el fin de impulsar la adaptación de la industria nacional al paradigma 4.0 y promover el desarrollo de soluciones tecnológicas 4.0 en el país, recuperando la experiencia internacional pero basándose en las particularidades del entramado productivo, con sus fortalezas y debilidades, para llevar adelante un proceso inteligente que mejore la competitividad de la economía y genere puestos de trabajo de calidad.

Este Plan fomentará la adopción y creación de soluciones 4.0, acompañando a las empresas que aún no han dado sus primeros pasos en este camino, a aquellas cuya incorporación de tecnologías 4.0 es reciente e incipiente, y a las que las han adoptado exitosamente. También trabajará con actores e instituciones que provean soluciones 4.0, y generará espacios de articulación entre oferentes y demandantes.

En este contexto, la madurez digital es un concepto crucial para que una organización o empresa pueda integrarse en la Industria 4.0. La madurez digital se puede definir como la alineación de las personas, la cultura, la estructura y las tareas para aprovechar las oportunidades que ofrece la infraestructura tecnológica, se centra en cómo las organizaciones pueden adaptarse y aplicar la tecnología digital para competir de manera eficaz en un entorno cada vez más digital [7, 8]. Es un mecanismo de diagnóstico organizacional moderno que le permite a las organizaciones definir estrategias de transformación digital orientadas a mejorar la competitividad y a desarrollar capacidades de adaptación y anticipación a los cambios del mercado y del entorno tecnológico. Las organizaciones digitalmente maduras tienen más probabilidades de ser tolerantes al riesgo, ágiles y capaces de impulsar el cambio digital [9].

En un modelo de medición se pueden identificar algunas de estas dimensiones: cultura digital, tecnología, procesos y operaciones, estrategia digital, organización, habilidades digitales, innovación, información y experiencia del cliente, gobernanza, visión, ecosistemas, liderazgo, regulación y seguridad, productos y servicios, modelo de negocio. Algunos de los modelos y herramientas que se destacan para diagnosticar o medir la madurez digital son: Modelo Acatech [10], Chequeo Digital [11], Kit PyMEs [12] y Metodología analítica descriptiva [13, 14].

Este proyecto de investigación se enmarca en la necesidad de evaluar el estado actual de la madurez digital de las PYMEs, especialmente en regiones como Jujuy, donde la transformación digital es crucial para la competitividad y el crecimiento económico. La

"Cámara PyME Jujuy"[1] tiene como objetivo que las empresas PyMEs asociadas adquieran un preponderante rol protagónico en el quehacer económico productivo de Jujuy, oportunamente se concretaron acciones relevantes para ampliar servicios que contribuyan a la mejora de la gestión empresarial; a la salud y seguridad de los empresarios, su grupo familiar y empleados; al intercambio seguro entre asociados y al asesoramiento para la resolución de problemas empresariales.

La propuesta de este proyecto consiste en realizar un estudio de las herramientas software que se utilizan en la actualidad para evaluar la madurez digital. A partir del estudio realizado, se prevé definir los factores y/o indicadores que son utilizados para la evaluación de las PyMEs, y de esta forma se podrá proporcionar un diagnóstico detallado que permita conocer, las fortalezas y debilidades en lo que respecta a sus capacidades tecnológicas, a partir de las cuales será posible realizar recomendaciones en el diseño de planes de acción concretos para la transformación digital de la PyME y de esta manera formar parte de la industria 4.0.

En la actualidad el Ministerio de Desarrollo Económico y Producción de Jujuy cuenta con un "Consultorio PyME" [15] disponible en su página web, que permite realizar un diagnóstico técnico y descriptivo gratuito, del estado de situación actual de una PyME, identificando problemas y debilidades, y sugiriendo oportunidades de mejora. La incorporación de una herramienta que permita evaluar la madurez digital de las PyMEs sería de gran utilidad para conocer su situación actual respecto a la industria 4.0, además, de aportar recomendaciones que le permitan la identificación de áreas de mejora y la planificación de estrategias de transformación digital adaptadas a sus capacidades y necesidades con el fin de que puedan formar parte de esta industria.

[1] https://camarapymejujuy.com.ar/

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El proyecto se encuentra adecuado a las líneas prioritarias y estratégicas acordadas por la Universidad Nacional de Jujuy según resolución C.S. N° 0110/23, en la que se incluye la línea de investigación correspondiente a "*Economía del Conocimiento*". Además, el proyecto corresponde a la modalidad Promover (P) la cual está dirigida a estimular las actividades de ciencia y tecnología de docentes-investigadores.

3. RESULTADOS OBTENIDOS/ESPERADOS

El proyecto de investigación tiene como objetivo estudiar las herramientas software que realizan la medición y diagnóstico de la madurez digital en pequeñas y medianas empresas (PyMEs), con el fin de evaluar la eficacia y aplicabilidad de estas herramientas en PyMEs de Jujuy. Además, de identificar y definir parámetros de evaluación específicos para proporcionar recomendaciones en su transformación digital hacia la industria 4.0.

El impacto esperado de este proyecto, es el siguiente:

- La identificación de factores específicos empleados en la evaluación de la madurez digital de las PyMEs de la provincia de Jujuy, lo que les permitirá conocer su situación actual respecto de las tecnologías asociadas a la industria 4.0.
- Los resultados obtenidos a partir de la evaluación de la madurez digital permitirán definir planes de acción adecuados para su transformación digital y posterior transición hacia la digitalización.
- La correcta evaluación de la madurez digital facilitará una transformación digital efectiva, permitiendo que las PyMEs puedan adaptarse a las tecnologías

digitales y mejoren su competitividad en el mercado.

4. FORMACIÓN DE RECURSOS HUMANOS

La estructura del equipo de investigación para la realización de este proyecto es la siguiente:

● Integrante Docente Investigador

-Directora: Mg. Ing. Nélida Raquel Cáceres. Categoría de Investigación IV. Experticia en CyT: Ingeniería de Software, Análisis y Diseño de Sistemas, Paradigma Orientado a Objetos, Bases de Datos, Data Mining, Data Streams, Realidad Aumentada.

-Codirectora: Ing. Ana Carolina Tolaba. Categoría de Investigación V. Actualmente realizando tesis de doctorado vinculada al área de modelado conceptual de datos a través de modelos semánticos. Experticia en CyT: Uso, desarrollo y optimización de metodologías para el modelado y diseño de aplicaciones de software. Técnicas y Herramientas de Soporte al Desarrollo de Sistemas de Información guiado por Ontologías. Web Ontology, Language (OWL), Semantic Web Rule Language (SWRL).

-Investigador: Mg. Ing. José Rolando Quispe. Docente de la Facultad de Ingeniería de la UNJu. Experticia en CyT: Análisis, diseño e implementación de software en Java. Manejo de base de datos MySQL y PostgreSQL. Seguridad en Sistemas y Bases de Datos.

● Integrante en Formación

- Mg. Lic. Liliana Mabel Medina. Docente de la Facultad de Ingeniería de la UNJu. Experticia en CyT: Especialista en Tecnología Multimedia para desarrollos educativos. Especialista en docencia en entornos digitales.

Con la realización de este proyecto de investigación se espera:

- La consolidación de los miembros del grupo.
- La incorporación de alumnos avanzados de las carreras afines de la facultad de ingeniería para su formación como investigadores.
- La dirección de trabajos finales de grado y participación en becas cuyas temáticas son propias del mencionado proyecto de investigación.

5. BIBLIOGRAFÍA

[1] del Val Román, J. L. (2016). "Industria 4.0: la transformación digital de la industria". Conferencia de Directores y Decanos de Ingeniería Informática, Informes CODDII. https://coddii.org/wp-content/uploads/2016/10/Informe-CODDII-Industria-4.0.pdf

[2] Ministerio de Trabajo, Empleo y Seguridad Social (2023). "El impacto de la Revolución 4.0 en el mercado laboral de las y los jóvenes en Argentina. Una perspectiva desde la sostenibilidad de la Seguridad Social". https://www.argentina.gob.ar/sites/default/files/3._el_impacto_de_la_revolucion_4.0_en_el_mercado_laboral_de_las_y_los_jovenes_en_argentina.pdf

[3] Basco, A. I., Beliz, G., Coatz, D., & Garnero, P. (2018). "Industria 4.0: fabricando el futuro" (Vol. 647). Inter-American Development Bank.

[4] Ministerio de Desarrollo Productivo Argentina. (2021). "Plan de Desarrollo Productivo Argentina 4.0". https://www.argentina.gob.ar/sites/default/files/plan_de_desarrollo_productivo_argentina_4.0.vf__2.pdf

[5] Albrieu, R., Basco, A. I., Brest López, C., De Azevedo, B., Peirano, F., Rapetti, M., & Vienni, G. (2019). "Travesía 4.0: hacia la transformación industrial argentina". BID, INTAL, CIPPEC y UIA.

[6] Nieponice, G., Rivera, R., Tfeli, A., & Drewanz, J. (2018). "Acelerando el desarrollo de la industria 4.0 en Argentina". Boston Consulting Group. Ciudad de Buenos Aires.

[7] Kane, G. C., Palmer, D., & Phillips, A. N. (2017). "Achieving digital maturity". MIT Sloan Management Review. 59 (1).

[8] Rader, D. (2019). “Digital maturity–the new competitive goal. Strategy & Leadership”, 47(5), 28-35. https://doi.org/10.1108/SL-06-2019-0084

[9] Lee, E., Barrett, M., Prince, K. & Oborn, E. “Digital maturity as process: A continuous approach to developing your digital maturity”. https://www.cambridgedigitalinnovation.org/single-post/digital-maturity-as-process-a-continuous-approach-to-developing-your-digital-maturity

[10] Schuh, G., Anderl, R., Gausemeier, J., Ten Hompel, M., & Wahlster, W. (2018). “Industrie 4.0 maturity index. Managing the digital transformation of companies”, 61. https://www.acatech.de/wp-content/uploads/2018/03/acatech_STUDIE_Maturity_Index_eng_WEB.pdf

[11] Del Do, A. M., Villagra, A., & Pandolfi, D. (2023). “Una Propuesta para la Transformación Digital en las PYMES”. RevITA, Revista Electrónica del Instituto de Tecnología Aplicada de la Universidad Nacional de la Patagonia Austral. 1(1), 7-21. https://publicaciones.unpa.edu.ar/index.php/RevITA/article/view/979/1047

[12] Landeta Etxeberria, A. (2019). “Kit pymes «estrategia digital»: Un modelo sintético para facilitar la elaboración del plan de evolución de la estrategia digital en el marco de la industria 4.0”. Revista de Contabilidad y Tributación, (435), 163-200. https://udimundus.udima.es/handle/20.500.12226/323

[13] Armijos-Buitrón, V. A., Espinoza-Vélez, M. P., & Rodríguez-Morales, G. “Nivel de madurez digital en micro, pequeñas y medianas empresas del sur de Ecuador: Estrategias para el fortalecimiento”. https://www.iiis.org/CDs2024/CD2024Spring/papers/CB714CK.pdf

[14] North, K., Aramburu, N., y Lorenzo, O. J. (2020). “Promoting digitally enabled growth in SMEs: a framework proposal”. Journal of Enterprise Information Management, 33(1), 1741- 0398

[15] CONSULTORIO PYME - Ministerio de Desarrollo Económico y Producción de Jujuy. (2022, agosto 3). Ministerio de Desarrollo Económico y Producción de Jujuy. http://www.produccion.jujuy.gob.ar/?page_id=3259

INGENIERÍA DE SOFTWARE EMPÍRICA: PROCEDIMIENTO PARA EL DESARROLLO Y MANTENIMIENTO DE COLECCIONES DE PROYECTOS DE SOFTWARE

Emanuel Irrazabal; Rubén Alfredo Bernal; Juan Andrés Carruthers; Andrea Lezcano Airaldi; Medina, Yanina, Romera Lourdes

Facultad de Ciencias Exactas y Naturales y Agrimensura.
Universidad Nacional del Nordeste
{eirrazabal, jacarruthers}@exa.unne.edu.ar

RESUMEN

Este es el primer año del proyecto F001-2021; una derivación de los proyectos F07-2009, F10-2013 y F018-2017, que desarrollaron modelos, métodos y herramientas para el desarrollo de aplicaciones software teniendo en cuenta aspectos de calidad de software. Este proyecto se enfoca en el diseño y desarrollo de estudios empíricos de calidad de software, como un insumo para la Ingeniería de Software basada en Evidencia.

La línea principal de investigación es la construcción de un modelo de proceso y procedimientos asociados para elaborar catálogos de proyectos software de calidad. Se atiende la necesidad de los grupos de investigación para obtener muestras curadas de proyectos imprescindibles para la generación de resultados confiables y generalizables en la experimentación de estudios empíricos de la calidad de software, proporcionando los insumos y procedimientos necesarios para conseguirlo de manera efectiva.

Otras líneas secundarias de investigación son el desarrollo de estudios empíricos para la toma de decisiones en grupo y el desarrollo de aplicaciones orientadas entornos científicos.

Palabras clave: calidad de software, estudios empíricos, código fuente.

CONTEXTO

Las líneas de Investigación y Desarrollo presentada en este trabajo corresponden al proyecto PI-21F001 “Desarrollo de estudios empíricos en Ingeniería del Software”, acreditado por la Secretaría de Ciencia y Técnica de la Universidad Nacional del Nordeste (UNNE) para el periodo 2022-2025, y a la beca interna doctoral de CONICET otorgada por RESOL-2021-154-APN-DIR#CONICET para el período 2021-2025.

1. INTRODUCCIÓN

La Ingeniería del Software Empírica (ISE) es una disciplina dedicada a la recopilación y análisis de datos derivados del desarrollo y mantenimiento de software. Su propósito es evaluar herramientas, prácticas y procedimientos que contribuyan, por ejemplo, a la corrección de errores, la identificación de vulnerabilidades o la estimación del tiempo de desarrollo. Todo esto con el objetivo de mejorar la calidad del software, basándose en evidencia científica obtenida a partir de estos estudios [1].

Dentro de la ISE, se utilizan diversos métodos de investigación con el fin de generar resultados replicables y generalizables, tales como experimentos, encuestas y estudios de minería de repositorios de software [2]. La generalización implica que los resultados

obtenidos puedan aplicarse a toda la población de interés, lo que requiere el uso de muestras representativas [3]. Por otro lado, la replicación busca validar hallazgos previos mediante la repetición de estudios bajo condiciones similares, esperando obtener resultados consistentes con los originales [4]. Para ello, es fundamental que los investigadores adopten estrategias de muestreo adecuadas y documenten detalladamente el protocolo del estudio.

Una fuente clave de datos en estudios empíricos es el código fuente de los proyectos de software [5]. En este contexto, plataformas como GitHub, GitLab o Bitbucket facilitan el acceso a información de millones de proyectos, lo que permite la realización de investigaciones en ISE. Sin embargo, no todos estos proyectos son adecuados para estudios científicos, por lo que su selección debe hacerse cuidadosamente [6]. Debido al esfuerzo que implica la recopilación de datos, algunos investigadores ponen a disposición de la comunidad científica las colecciones de proyectos utilizadas en sus estudios.

Aun así, estudios secundarios [7, 8] han identificado ciertas prácticas en la literatura que pueden afectar la validez de los resultados obtenidos con estas colecciones de datos. Entre los problemas metodológicos detectados se encuentran el uso de muestras reducidas, la ausencia de criterios explícitos para evaluar la calidad de los proyectos y la aplicación de técnicas de muestreo no probabilísticas. Además, también se han señalado deficiencias en la difusión de los instrumentos de investigación, como instrucciones incompletas, restricciones en el acceso a los conjuntos de datos y la falta de paquetes de replicación, entre otros aspectos.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

A continuación, se describe la línea de investigación y desarrollo, la cual fue tomada como prioritaria a partir de las necesidades obtenidas del estudio de las pymes y grupos de trabajo de la región [9]:

Curaduría de proyectos

Para abordar las necesidades mencionadas en la sección anterior, se propuso la construcción del modelo de proceso SUM4SOFT. El modelo detalla los pasos y reglas para construir colecciones representativas, y además ofrece plantillas para estandarizar la generación de los instrumentos de recolección de datos facilitando la replicación de los estudios. Para diseñar, construir y validar SUM4SOFT se empleó la metodología ciencia de diseño [10] compuesta de cinco actividades: explicar el problema, definir objetivo y requisitos, diseñar y desarrollar el artefacto, demostrar el artefacto y evaluar el artefacto. Para la construcción de las actividades de esta metodología el grupo participó en un Simposio Doctoral para terminar de elaborar el plan de investigación [11].

3. RESULTADOS OBTENIDOS/ESPERADOS

En el marco de este proyecto y respecto de la línea de curaduría de proyectos y de acuerdo con lo desarrollado en el año 2024 se logró la mejora y utilización de herramientas para dar soporte en la generación de métricas obtenidas del análisis estático del código fuente [12].

La contribución principal de esta investigación fue el desarrollo de herramientas para el soporte de estudios en la Ingeniería del Software Empírica. En primer lugar, el modelo de proceso para la construcción y actualización de muestras de proyectos software SUM4SOFT. En la figura 1 se muestra un recorte de la perspectiva de comportamiento del modelo.

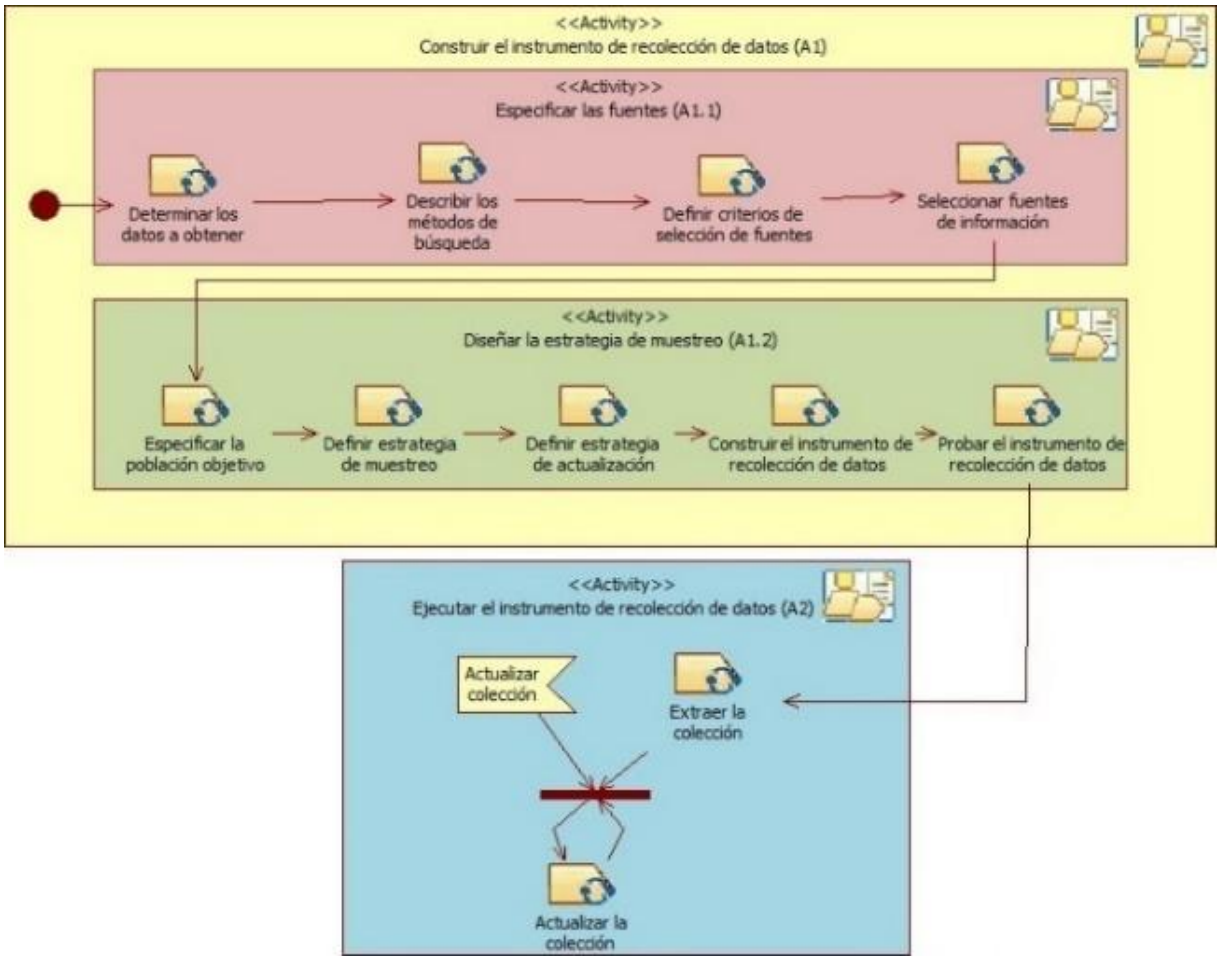


Figura 1. Perspectiva de comportamiento de SUM4SOFT.

En segundo lugar, las salidas de SUM4SOFT obtenidos en el contexto de un estudio de minería de repositorios: una colección de proyectos software y el instrumento de recolección de datos JavaQ con las herramientas para mantener la colección facilitando la replicación de los resultados.

También se profundizó acerca del estado actual de las colecciones de proyectos software utilizadas en la ISE a través de tres estudios empíricos: dos mapeos sistemáticos [13][14] y un estudio longitudinal [15]. Las revisiones de la literatura confirmaron algunas problemáticas mencionadas por otros autores como: la falta de estrategias de selección de proyectos estandarizadas, la omisión de la vigencia como criterio de selección, la ausencia del muestreo probabilístico y las restricciones de acceso, tanto a las colecciones como a las herramientas empleadas en la experimentación con ellas. A su vez identificaron las características deseables en los proyectos software para construir una colección curada. Por su parte, el estudio longitudinal midió el impacto del tiempo en una población de proyectos y demostró la importancia de revisar la vigencia de las colecciones en contextos donde la población evoluciona de manera constante.

El objetivo del estudio de minería de repositorios fue evaluar la efectividad de JavaQ para construir y actualizar colecciones de proyectos. Para ello, se generaron los umbrales de calidad de tres métricas de código: complejidad ciclomática (McCC), cantidad de líneas de código (LOC) y peso de métodos por clase (WMC), habiendo analizado previamente la importancia de estas métricas en la mantenibilidad del código fuente [16]. Los

umbrales de métricas delimitan puntos de referencia aceptables y anormales en los componentes de software, proporcionando un indicador para evaluar el estado del sistema, utilizando herramientas de libres [17].

4. FORMACIÓN DE RECURSOS HUMANOS

En esta línea de trabajo del Grupo de Investigación sobre Calidad de Software (GICS) están involucrados 3 docentes investigadores, dos becarios internos doctorales de CONICET, uno de ellos trabajando en el desarrollo del modelo y el segundo becario apoyando en la construcción de la experimentación.

5. REFERENCIAS

[1] B. Kitchenham, T. Dyba, and M. Jorgensen, “Evidence-based software engineering,” in Proceedings. 26th International Conference on Software Engineering, IEEE Comput. Soc, 2004, pp. 273–281. doi: 10.1109/ICSE.2004.1317449.

[2] C. Wohlin and A. Aurum, “Towards a decision-making structure for selecting a research design in empirical software engineering,” Empir Softw Eng, vol. 20, no. 6, pp. 1427–1455, Dec. 2015, doi: 10.1007/s10664-014-9319-7.

[3] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén, Experimentation in Software Engineering, vol. 9783642290. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. doi: 10.1007/978-3-642-29044-2.

[4] E. Barr, J. Bell, M. Hilton, S. Mechtaev, and C. Timperley, “Continuously Accelerating Research,” in 2023 IEEE/ACM 45th International Conference on Software Engineering: New Ideas and Emerging Results (ICSE-NIER), IEEE, May 2023, pp. 123–128. doi: 10.1109/ICSE-NIER58687.2023.00028.

[5] M. M. Lehman, “Laws of software evolution revisited,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer Verlag, 1996, pp. 108–124. doi: 10.1007/BFb0017737.

[6] N. Munaiah, S. Kroh, C. Cabrey, and M. Nagappan, “Curating GitHub for engineered software projects,” Empir Softw Eng, vol. 22, no. 6, pp. 3219–3253, Dec. 2017, doi: 10.1007/s10664-017-9512-6.

[7] S. Baltes and P. Ralph, “Sampling in software engineering research: a critical review and guidelines,” Empir Softw Eng, vol. 27, no. 4, p. 94, Jul. 2022, doi: 10.1007/s10664-021-10072-8.

[8] V. Cosentino, J. Luis, and J. Cabot, “Findings from GitHub,” in Proceedings of the 13th International Conference on Mining Software Repositories, New York, NY, USA: ACM, May 2016, pp. 137–141. doi: 10.1145/2901739.2901776.

[9] Dapozo, Gladys N., et al. "Características del desarrollo de software en la ciudad de Corrientes." *XXI Congreso Argentino de Ciencias de la Computación (Junín, 2015)*. 2015.

[10] P. Johannesson and E. Perjons, An Introduction to Design Science, vol. 9783319106. Cham: Springer International Publishing, 2014. doi: 10.1007/978-3-319-10632-8.

[11] Pinto Oppido, J. A., Carruthers, J. A., & Irrazábal, E. A. (2021). Sonar JUploader, aplicación para el análisis, sincronización y actualización automática de proyectos a Sonar Cloud. In IV Jornadas de Calidad de Software y Agilidad.

[12] Carruthers, J. A. (2022). Open-Source Software Projects Curating Model for Empirical Software Engineering Studies. Anais Do XXV Congresso Ibero-Americano Em Engenharia de Software (CIbSE 2022), 416–423. https://doi.org/10.5753/cibse.2022.20992

[13] Carruthers, J. A., Diaz-Pace, J. A., & Irrazabal, E. A. (2022). How are software datasets constructed in Empirical Software Engineering studies? A systematic mapping study. 2022 48th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), 442–450. https://doi.org/10.1109/SEAA56994.2022.00075

[14] Carruthers, J. A., Diaz-Pace, J. A., & Irrazabal, E. A. (2022). A Systematic Mapping Study of Empirical Studies Performed with Collections of Software Projects. Computación y Sistemas, 26(4). https://doi.org/10.13053/cys-26-4-3981.

[15] Carruthers, J. A., Diaz-Pace, J. A., & Irrazábal, E. (2024). A longitudinal study on the temporal validity of software samples. Information and Software Technology, 168, 107404. https://doi.org/10.1016/j.infsof.2024.107404.

[16] Irrazábal, Emanuel, and Javier Garzás. "Análisis de métricas básicas y herramientas de código libre para medir la mantenibilidad." *Reicis. Revista española de innovación, calidad e ingeniería del software* 6.3 (2010): 56-65.

[17] Irrazábal, E., Garzás, J., & Marcos, E. (2011). Alignment of Open Source Tools with the New ISO 25010 Standard-Focus on Maintainability. In International Conference on Software and Data Technologies (Vol. 2, pp. 111-116).

INTERACCIÓN POR VOZ AFECTIVA Y SEGURA: UN ENFOQUE HÍBRIDO DE PLN, LÓGICA DIFUSA Y BLOCKCHAIN

Fransoy, Marcelo[1]; Hossian, Alejandro[1]; Merlino, Hernán[1];Pollo-Cattaneo, María F.[1]
[1] Grupo de Estudio en Metodologías de Ingeniería en Software (GEMIS)
Departamento de Sistemas de Información (GEMIS)
Universidad Tecnológica Nacional. Facultad Regional Buenos Aires.
marcelo.fransoy@gmail.com, alejandrohossian@yahoo.com.ar, hmerlino@frba.utn.edu.ar, flo.pollo@gmail.com

RESUMEN

La interacción afectiva entre humanos y máquinas es un campo en auge gracias a los sistemas de interacción por voz. Estos sistemas buscan transformar la comunicación con la tecnología, pero para lograrlo, es crucial que comprendan y respondan a las emociones humanas de manera natural y empática.

El procesamiento del lenguaje natural (PLN) es una herramienta clave para analizar el lenguaje y extraer información relevante para la detección de emociones. La lógica difusa, por su parte, permite abordar la complejidad de las emociones humanas, representando y clasificando emociones incluso en situaciones ambiguas. Además, la tecnología Blockchain promete garantizar el almacenamiento seguro de datos emocionales.

Este trabajo se centra en la combinación de PLN, lógica difusa y Blockchain para crear sistemas de interacción por voz más avanzados y seguros. El objetivo principal es diseñar un sistema capaz de reconocer y responder a las emociones del usuario en español de manera precisa, personalizada y segura. Se espera que este sistema contribuya al campo de la interacción por voz y la Inteligencia Artificial (IA) emocional, abriendo nuevas vías para asistentes virtuales y chatbots más empáticos y personalizados.

Palabras clave: interacción por voz, PLN, lógica difusa, blockchain, emociones.

CONTEXTO

La Universidad Tecnológica Nacional ha definido diversos Programas de Investigación, Desarrollo e Innovación, entre los cuales se encuentra el Programa de Sistemas de Información e Informática que tiene como objetivo “intensificar y focalizar las acciones tendientes a fortalecer y promover el crecimiento de temáticas de investigación en Sistemas de Información e Informática, y promover la interacción con la industria informática en general y de desarrollo de sistemas de información y de software en particular”[1].

En la Facultad Regional Buenos Aires de la Universidad Tecnológica Nacional, desarrolla sus actividades, entre otros grupos de investigación, el grupo GEMIS.BA aprobado por la ordenanza N° 1292 de la Universidad Tecnológica Nacional y la resolución N° 3245/2023 del Consejo Directivo de la Facultad Regional Buenos Aires, el cual trabaja en el campo de la Ingeniería en Sistemas de Información, Ingeniería de Software e Ingeniería del Conocimiento.

El Grupo GEMIS.BA realiza, en el marco de las temáticas objeto de este trabajo, diversos proyectos asociados a la línea de la IA con proyectos de investigación y desarrollo y actividades de docencia e investigación en diferentes instituciones contando con trabajos finales de especialización y tesis de maestría.

1. INTRODUCCIÓN

La interacción afectiva entre humanos y máquinas se ha convertido en un campo de investigación de gran relevancia en la era digital [2]. A medida que los sistemas de interacción por voz se vuelven más comunes en nuestra vida cotidiana, surge la necesidad de

desarrollar sistemas que no sólo comprendan el lenguaje humano, sino que también sean capaces de interpretar y responder a las emociones de manera natural y empática [3][4].

En este contexto, el PLN se ha consolidado como una herramienta fundamental para analizar el lenguaje natural y extraer información semántica y contextual relevante para la detección de emociones. El PLN permite identificar y analizar matices emocionales en diversos idiomas y contextos, sentando las bases para un sistema de interacción por voz más preciso y adaptable [5][6].

Sin embargo, las emociones humanas son complejas y subjetivas, lo que plantea desafíos para su detección y modelado. La lógica difusa ofrece un marco flexible y robusto para abordar esta complejidad, permitiendo representar y clasificar emociones incluso en situaciones de ambigüedad y superposición de estados emocionales [7][8]. La integración del PLN con la lógica difusa permite aprovechar las ventajas de ambos enfoques, logrando una comprensión más completa y precisa de las emociones expresadas a través del lenguaje natural [9].

Además, la creciente preocupación por la privacidad y la seguridad de los datos emocionales exige soluciones innovadoras que garanticen la protección y el control de la información sensible. La tecnología Blockchain, con sus propiedades de descentralización, transparencia e inmutabilidad, se presenta como una solución prometedora para abordar estos desafíos. Al utilizar Blockchain, se puede garantizar el almacenamiento y el intercambio seguro de datos emocionales, brindando a los usuarios un mayor control sobre su información y fomentando la confianza en los sistemas de interacción por voz [10].

En este contexto, resulta fundamental explorar cómo la combinación de PLN, lógica difusa y Blockchain puede dar lugar a sistemas de interacción por voz más avanzados y seguros. Estos sistemas podrían mejorar la precisión en la detección de emociones, generar respuestas afectivas personalizadas y garantizar la privacidad de los datos emocionales del usuario.

Además de estas tecnologías, la investigación en interacción afectiva también destaca la importancia de considerar el contexto conversacional, las modalidades múltiples (voz, texto, expresiones faciales) y la adaptabilidad a diferentes culturas e industrias [11]. Un enfoque holístico que tenga en cuenta todos estos factores es esencial para lograr una interacción más natural, empática y efectiva [12][13].

En resumen, la interacción afectiva entre humanos y máquinas es un campo de investigación en constante evolución, con el objetivo de crear sistemas que sean capaces de comprender y responder a las emociones humanas de manera natural y empática [14][15][16][17]. El PLN, la lógica difusa y la tecnología Blockchain se presentan como herramientas clave para superar los desafíos actuales en la detección de emociones, la generación de respuestas afectivas y la protección de la privacidad en los sistemas de interacción por voz.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

La Universidad Tecnológica Nacional (UTN) de Argentina ha demostrado un firme compromiso con el avance de la investigación científica a través de la implementación de diversos programas de Investigación, Desarrollo e Innovación. Un ejemplo destacado es el Programa de Sistemas de Información e Informática, establecido en 2016 mediante la Resolución 2508/16. Este programa tiene como objetivo principal fortalecer y expandir las actividades de investigación en áreas clave como los Sistemas de Información y la Informática, fomentando al mismo tiempo una colaboración más

estrecha con la industria informática, especialmente en los sectores de desarrollo de software y sistemas de información.

En este contexto, y como parte de una iniciativa más amplia para impulsar la investigación y el desarrollo en Ingeniería en Sistemas de Información e Ingeniería del Software, se iniciaron en 2009 las actividades del Grupo de Estudio e Investigación en Modelos y Sistemas de Información (GEMIS), grupo que queda formalizado y aprobada como GEMIS.BA por la ordenanza N° 1292 de la Universidad Tecnológica Nacional y la resolución N° 3245/2023 del Consejo Directivo de la Facultad Regional Buenos Aires.

Ubicado en el Departamento de Ingeniería en Sistemas de Información de la Facultad Regional Buenos Aires de la UTN (UTN-FRBA), GEMIS.BA reúne a un equipo multidisciplinario de docentes, graduados y estudiantes dedicados a la sistematización y difusión del conocimiento. Su trabajo abarca una amplia gama de aplicaciones y metodologías, lo que refleja el compromiso de la UTN con la investigación aplicada y la colaboración entre la academia y la industria. La creación de GEMIS.BA es un claro ejemplo del enfoque de la universidad hacia la innovación tecnológica y su aplicación en la sociedad.

En el ámbito de GEMIS.BA, se han llevado a cabo diversas investigaciones relacionadas con la IA. Una de las primeras iniciativas fue el Proyecto de Investigación y Desarrollo (PID) titulado "Implementación de sistemas inteligentes para la asistencia a alumnos y docentes de la carrera de ingeniería en sistemas de información", que se desarrolló entre 2016 y 2019. Este proyecto se centró en la aplicación de tecnologías de IA y desarrollos de software inteligente para abordar desafíos en el ámbito educativo.

Posteriormente, entre 2019 y 2022, se implementó otro PID denominado "Prácticas ingenieriles aplicadas para la implementación de Sistemas Inteligentes basados en Machine Learning". En este caso, el proyecto se enfocó en analizar el uso de buenas prácticas de ingeniería para el desarrollo de software inteligente que utiliza Machine Learning.

A partir de 2023, se inició el proyecto "Inteligencia Artificial para el análisis predictivo en salud mental", cuyo objetivo es proponer técnicas, tareas y métodos para la implementación de Modelos Predictivos basados en IA en el campo de la salud mental. Además, se planea incorporar el análisis de señales de EEG como complemento a los modelos predictivos desde 2024, aprovechando la información neurofisiológica proporcionada por el EEG.

3. RESULTADOS OBTENIDOS/ESPERADOS

El resultado esperado de este trabajo es diseñar e implementar un sistema de interacción por voz en español para asistentes virtuales y chatbots que sea capaz de reconocer y responder a las emociones del usuario de manera apropiada y personalizada, garantizando su seguridad y privacidad mediante la tecnología blockchain.

En la exploración inicial de esta línea de investigación, se ha realizado un exhaustivo análisis del estado del arte en el campo PLN, la lógica difusa, el reconocimiento de emociones en la voz y la tecnología blockchain. Esto ha proporcionado una comprensión profunda del panorama actual y las tendencias en estas áreas, identificando oportunidades y desafíos para el desarrollo del sistema propuesto.

A partir de este conocimiento, se espera obtener los siguientes resultados específicos:

Desarrollo de un modelo híbrido PLN-Lógica Difusa: Se espera desarrollar un modelo híbrido que combine técnicas de PLN y lógica difusa para mejorar la precisión en la detección y clasificación de emociones en la voz del usuario. Este modelo considerará contextos conversacionales y sutilezas emocionales,

superando las limitaciones de los enfoques tradicionales.

Implementación de un mecanismo de generación de respuestas emocionales: Se espera implementar un mecanismo que utilice la información emocional detectada para generar respuestas personalizadas, empáticas y contextualmente relevantes. Este mecanismo adaptará las respuestas al estado emocional del usuario, mejorando la calidad y la personalización de la interacción.

Diseño e integración de una arquitectura blockchain: Se espera diseñar e integrar una arquitectura blockchain que permita el almacenamiento seguro y transparente de los datos emocionales del usuario. Esta arquitectura garantizará la privacidad y el control del usuario sobre su información, abordando preocupaciones éticas y de seguridad.

Evaluación del sistema: Se espera evaluar el sistema propuesto mediante pruebas de usuario y métricas objetivas. Estas evaluaciones medirán la precisión en la detección de emociones, la calidad de las respuestas emocionales generadas y la percepción de seguridad y privacidad por parte de los usuarios.

Análisis del impacto en la experiencia del usuario: Se espera analizar el impacto del sistema en la experiencia del usuario y la calidad de la interacción con asistentes virtuales y chatbots en diferentes aplicaciones y contextos. Este análisis proporcionará información valiosa sobre la efectividad y el potencial del sistema para mejorar la interacción humano-máquina.

Se espera que los resultados de este trabajo contribuyan significativamente al campo de la interacción por voz y la IA emocional, abriendo nuevas vías para el desarrollo de asistentes virtuales y chatbots más empáticos, personalizados y seguros. Además, se espera que este trabajo siente las bases para futuras investigaciones en el área de la interacción humano-máquina y la aplicación de la tecnología blockchain en este contexto.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de investigación para este trabajo se encuentra conformado por investigadores formados, investigadores de apoyo y alumnos de grado y posgrado. Esta línea de trabajo busca tanto la obtención de nuevos conocimientos como la motivación de los implicados para su desarrollo en la carrera de investigadores y profesionales, además de fomentar la aplicación de las temáticas en su propia actividad profesional.

En el marco del proyecto de investigación sobre el diseño e implementación de un sistema de interacción por voz en español para asistentes virtuales y chatbots con reconocimiento de emociones, se ha puesto especial énfasis en la formación de recursos humanos, integrando a estudiantes y docentes en diversas etapas de su desarrollo profesional y académico. Este enfoque multidisciplinario permitirá que varios estudiantes de grado y posgrado participen activamente en el proyecto, desarrollando sus tesis y proyectos de investigación en áreas clave como el PLN, la lógica difusa, el reconocimiento de emociones en la voz y la tecnología blockchain.

Se prevé que las investigaciones vinculadas al Proyecto de tesis de Magíster se desarrollen en GEMIS.BA, a cargo de la Dra. María Florencia Pollo Cattaneo, en el marco del Proyecto Inteligencia Artificial para el Análisis Predictivo en Salud Mental, acreditado por la Universidad Tecnológica Nacional, Facultad Regional Buenos Aires.

Asimismo, GEMIS.BA tiene vínculos con otros grupos de investigación que trabajan en áreas vinculadas a las temáticas de tesis. Estas vinculaciones permitirán intercambiar opiniones con tesistas de posgrado e investigadores que trabajen en líneas de investigación afines.

Se espera que esta línea de trabajo permita:

- La generación de nuevas oportunidades y experiencias para la formación de los investigadores.
- El desarrollo de por lo menos 2 tesis de Maestría en Ingeniería en Sistemas de Información.
- La participación de becarios graduados a través de las becas BINID.

5. BIBLIOGRAFÍA

[1] Universidad Tecnológica Nacional (2016). Resolución Nro. 2508/16

[2] Yashwanth, V., Sumalatha, M. R., Yukan, A. S., Narayan, A. B., & Paliyela, L. H. (2022, December). Real-time Speech Emotion Detection using Artificial Intelligence. In 2022 IEEE International Conference on Current Development in Engineering and Technology (CCET) (pp. 1-5). IEEE

[3] Roshan, R., Sohan, M. H., Raj, S. S., & Prasad, V. B. (2024, May). Sentient Sound waves: Elevating Emotional Communication with AI-Generated Speech Technology. In 2024 5th International Conference for Emerging Technology (INCET) (pp. 1-6). IEEE.

[4] Srikanth, S. S., Manikantan, S., Posonia, A. M., Mary, S. P., & Revathy, S. (2023, August). Voice Assistant with Emotion Recognition. In 2023 International Conference on Circuit Power and Computing Technologies (ICCPCT) (pp. 799-804). IEEE.

[5] Suresh, C., Sathvik, M. C., Deepthi, N., Purnima, K. M. S., & Chouhan, K. P. S. (2023, March). A study on cross-lingual speech emotion analysis using natural language processing. In 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS) (pp. 808-815). IEEE.

[6] Tseng, S. Y., Narayanan, S., & Georgiou, P. (2021). Multimodal embeddings from language models for emotion recognition in the wild. IEEE Signal Processing Letters, 28, 608-612.

[7] Jayashree, D., Pandithurai, O., Prasad, S., Suresh, A. S., & Vigneshwaran, S. (2021, October). Sentimental analysis on voice based reviews using fuzzy logic. In 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA) (pp. 1-6). IEEE.

[8] Abhishek, K., & Appuhamy, E. J. G. S. (2023, February). Emotion Classification Comparison in Convolutional Neural Networks and Fuzzy Logics for Service Robotic Applications. In 2023 3rd International Conference on Advanced Research in Computing (ICARC) (pp. 36-41). IEEE.

[9] Karthiga, G., Jayakarthika, K., Ilakkiya, N., Lavanyagayathri, V., & SD, U. M. (2023, November). A Fuzzy Logic and NLP Approach to Emotion Driven Response Generation for Voice Interaction. In 2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE) (pp. 1-5). IEEE.

[10] Sivasathiya, M. G., AR, H. R., & Kanishkaa, R. (2024, January). Emotion-Aware Multimedia Synthesis: A Generative AI Framework for Personalized Content Generation based on User Sentiment Analysis. In 2024 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT) (pp. 1344-1350). IEEE.

[11] Ganesh, D. B., Prasanth, B., Chowdary, M. T., Avala, V. S., & Gangashetty, S. V. (2024, January). Emotion Speech Recognition in the Cloud. In 2024 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT) (pp. 1033-1039). IEEE.

[12] Ding, J., Chen, X., Lu, P., Yang, Z., Li, X., & Du, Y. (2023). DialogueINAB: an interaction neural network based on attitudes and behaviors of interlocutors for dialogue emotion recognition. The Journal of Supercomputing, 79(18), 20481-20514.

[13] Carvalho, I., Oliveira, H. G., & Silva, C. (2023). The importance of context for sentiment analysis in dialogues. IEEE Access, 11, 86088-86103.

[14] Sivasathiya, M. G., AR, H. R., & Kanishkaa, R. (2024, January). Emotion-Aware Multimedia Synthesis: A Generative AI Framework for Personalized Content Generation based on User Sentiment Analysis. In 2024 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT) (pp. 1344-1350). IEEE.

[15] Dehbozorgi, N., & Kunuku, M. T. (2023). Exploring the Influence of Emotional States in Peer Interactions on Students' Academic Performance. IEEE Transactions on Education.

[16] Asha, P., Suresh, L. P., Gani, A. M. I., Praveen, S., Sai, A. D., & Mathivanan, G. (2024, April). Employing Intelligence for Detecting the Emotions Using Efficient Machine Learning Algorithms. In 2024 International Conference on Science Technology Engineering and Management (ICSTEM) (pp. 1-5). IEEE.

[17] Leu, F. Y., Liu, J. C., Hsu, Y. T., & Huang, Y. L. (2014). The simulation of an emotional robot implemented with fuzzy logic. Soft Computing, 18, 1729-1743.

INTRODUCCIÓN A LA COMPUTACIÓN CUÁNTICA Y ASPECTOS DE INGENIERÍA DE SOFTWARE Y BASES DE DATOS NOSQL PARA EL DESARROLLO DE SISTEMAS DE SOFTWARE HÍBRIDOS.

Luciano Marrero, *Pablo Thomas*, *Ariel Pasini*, *Rodolfo Bertone*, *Eduardo Ibáñez*, *Verena Olsowy*, *Leonardo Corbalán*, *Juan Fernández Sosa*, *Fernando Tesone*, *Patricia Pesado*

Instituto de Investigación en Informática LIDI (III-LIDI)
Facultad de Informática – Universidad Nacional de La Plata
50 y 120 La Plata Buenos Aires
Centro Asociado CIC
{lmarrero, pthomas, apasini, pbertone, eibanez, volsowy, corbalan, jfernandez, ftesone, ppesado}@lidi.info.unlp.edu.ar

RESUMEN

Se presenta una línea de investigación que tiene por objeto estudiar las problemáticas que afronta la Ingeniería de Software y las Bases de Datos (Relacionales y No Relacionales) ante una era digital que crece de forma constante. Los avances tecnológicos no se detienen y cada vez son más las personas que pueden acceder a un mundo gobernado por internet y sus diferentes aplicaciones. Además, en computación binaria, se está llegando a un límite en la miniaturización de los circuitos integrados experimentando problemas asociados a la física cuántica, como por ejemplo el "efecto túnel" (electrones se escapan de los canales por donde deben circular). Por lo tanto, la computación tradicional se encuentra limitada porque ha alcanzado escalas difíciles de seguir reduciendo [30, 34, 35].

En los últimos años, las investigaciones en tecnología cuántica han crecido exponencialmente. Diversas organizaciones y empresas invierten tiempo y dinero en el desarrollo de computadoras cuánticas. La computación cuántica representa un cambio radical en todos los aspectos para la computación binaria.

En este contexto, se deben afrontar diversas problemáticas tecnológicas; por un lado el continuo crecimiento y generación de datos que existen en la web y sus aplicaciones, y por otro lado, estudiar el impacto que podría producir el éxito de la computación cuántica en la profesión Informática tal cual se la conoce [36].

En el contexto de las Bases de Datos, la idea de considerar que un único modelo de datos pueda adaptarse de forma eficiente a todos los requerimientos es impensada. Se plantean diversas alternativas que permiten almacenar la información de forma no estructurada (Bases de Datos NoSQL) [1, 2, 36]

Además, es necesario nuevas formas de pensar el diseño de Bases de Datos que se adapte

a esta nueva dinámica del desarrollo del Software [2, 6, 8].

Por último, el avance de la tecnología ha impulsado además aspectos que hace algunos años no eran considerados en el desarrollo del Software, tales como, movilidad, geolocalización y la diversidad de los dispositivos electrónicos involucrados [5].

Palabras claves: Diseño no estructurado de datos, Procesos de diseño de Bases de Datos Relacionales y Bases de Datos NoSQL, Bases de Datos como servicio en la nube, Computación Cuántica.

CONTEXTO

Esta línea de Investigación forma parte del Proyecto "*Diseño, desarrollo y evaluación de sistemas en escenarios híbridos para áreas clave de la sociedad actual: educación, ciudades inteligentes y gobernanza digital*" del Instituto de Investigación en Informática LIDI de la Facultad de Informática, acreditado por el Ministerio de Educación de la Nación. Hay cooperación con Universidades de Argentina y se está trabajando con Universidades de Europa en proyectos financiados por el Ministerio de Ciencia y Tecnología de España y la AECID. Se utilizan los recursos de Hardware y Software disponibles en el III-LIDI para diseñar, desarrollar y probar diferentes soluciones a problemáticas relacionadas con escenarios a investigar. Como resultado de esto, se espera obtener métricas reales que sirvan como referencia para los investigadores en la comparación de resultados. Las publicaciones científicas generadas y la transferencia continua de resultados concretos validan esta línea de investigación.

1. INTRODUCCIÓN

"*Diseño, desarrollo y evaluación de sistemas en escenarios híbridos para áreas clave de la sociedad actual: educación, ciudades inteligentes y gobernanza digital*" del Programa de Incentivos, es un proyecto que propone profundizar las investigaciones que se vienen realizando en el III-LIDI y extender la mirada a nuevos desafíos y cambios que están en gestación. Se organiza en tres subproyectos que permiten atender de manera ordenada el objetivo general propuesto. "*SP1: Ingeniería de Software para escenarios híbridos*", "*SP2: Ciudades Inteligentes sostenibles (CIS). Gobernanza Digital. Buenas prácticas y calidad*" y "*SP3: Creación de tecnologías digitales para el escenario educativo".* [22].

Este artículo se centra en el subproyecto "*SP1 - Ingeniería de Software para escenarios híbridos*".

Se orienta a la investigación de metodologías y técnicas de la Ingeniería de Software, Bases de Datos (Relacionales y No Relacionales), estudio de procesos de diseño de Base de Datos No Relacionales, escenarios híbridos, aplicaciones móviles, computación cuántica, entre otras temáticas de interés [22, 35, 36].

Actualmente, en la ciencia informática se presentan un conjunto de nuevos desafíos que se deben afrontar de forma vertiginosa. Estos desafíos van desde cualquier etapa o nivel inicial hasta que se llega a un resultado o producto final.

En lo que respecta a las Bases de Datos, principalmente, las Bases de Datos NoSQL, en ciertas ocasiones, se ha pasado de tener que

pensar en un diseño o esquema detallado en la representación de sus datos a tener que plantear directamente las estructura a nivel físico teniendo en cuenta el problema a resolver, y la flexibilidad que poseen los esquemas no estructurados de datos [32, 33, 36].

Actualmente, no es común encontrar procesos de diseño establecidos para documentar la semántica en una Base de Datos NoSQL. No obstante, desde hace muchos años, es frecuente en la disciplina informática, agilizar los tiempos de las tareas a realizar a través de suprimir documentación que avale la toma de decisiones.

Es importante remarcar que el diseño de una Base de Datos es un proceso cuyo objetivo es definir la estructura adecuada para el almacenamiento de datos de un sistema de información. El diseño influye de forma directa en la eficiencia para almacenar, recuperar y analizar los datos en la Base de Datos.

Para las Bases de Datos relacionales, se cuenta con un proceso de diseño adoptado el cual inicia con un modelo conceptual, luego se deriva a un modelo lógico relacional y finaliza con un modelo físico de datos [6, 7].

Para el caso de las Bases de Datos NoSQL, no se cuenta un proceso masivamente aceptado para el diseño de esquemas de datos, sino que han surgido como soluciones directamente a nivel físico. No obstante, a medida que han adquirido popularidad, se han propuesto diferentes procesos y/o metodologías de diseño NoSQL, necesarios para comprender la semántica de los datos [28, 31].

Actualmente, muchas empresas desarrollan aplicaciones que son mundialmente utilizadas poniendo el foco en la escalabilidad de sus recursos y enfrentando nuevos retos para satisfacer la demanda de sus usuarios [2, 14, 20].

Existe un amplio abanico de tecnologías, tanto en Bases de Datos (Relacionales y No Relacionales), como así también, en lenguajes para el desarrollo de todo tipo de aplicaciones de Software [10, 11, 12].

En el contexto de Base de Datos, existen distintos motores de Bases de Datos Relacionales y No Relacionales para el desarrollo de aplicaciones de propósito general, entre ellos, podemos encontrar a SQLite [16], Couchbase [17], Firebase Realtime Database [23], Google Cloud Firestore [23], Redis [24], MongoDB [25], Cassandra [26], Neo4j [27], entre otras.

Los motores de Bases de Datos son parte de la continua evolución tecnológica, esto genera que el usuario tenga un gran abanico de posibilidades para elegir el motor de Base de Datos que se adapte a sus exigencias [1, 8, 9].

En la actualidad, se presenta la posibilidad de tener un cambio radical en la disciplina informática tal cual se la conoce. Este cambio viene impulsado por la computación cuántica.

La computación cuántica no es un tema actual, proviene de los años 80, y en esos años su estudio y crecimiento ha sido relativamente lento en comparación con otros avances tecnológicos y además, eran los profesionales de la física cuántica quienes más se centraban en el tema, pero desde hace varios años, el panorama es otro y ha cambiado de forma vertiginosa. Importantes empresas y organizaciones invierten tiempo y dinero en una competencia por el desarrollo de computadoras cuánticas [30, 35, 37].

Actualmente, existen lenguajes de programación cuántica diseñados específicamente para escribir algoritmos ejecutables en computadoras cuánticas. Este tipo

de lenguajes de programación opera con *qubits* o bit cuántico (análogo al *bit* en computación binaria) y aprovecha fenómenos como la superposición, el entrelazamiento y la interferencia (conceptos existentes en la física cuántica). Los algoritmos cuánticos implican la programación de circuitos cuánticos mediante la relación de puertas cuánticas (análogo a las compuertas lógicas de la computación binaria). Algunos ejemplos de estos lenguajes son: Qiskit de IBM, Cirq de Google, Q# de Microsoft, entre otros [35, 37, 38, 39].

Además, la computación cuántica en la nube o Quantum Cloud Computing (QCC) y los simuladores cuánticos en línea son dos herramientas que están democratizando el acceso a la computación cuántica en el mundo y abren un abanico de posibilidades para la investigación, el desarrollo y la innovación en diversos campos. Entre ellos se encuentran: IBM Quantum Composer de IBM, Google Quantum Playground de Google, Amazon Braket de Amazon, entre otros [37, 38, 39, 40].

Aunque la computación cuántica promete superar las prestaciones de las computadoras clásicas, la implementación de sistemas cuánticos a gran escala aún presenta desafíos técnicos y económicos significativos.

Otra temática de investigación se obtiene de analizar el proceso de despliegue del Software. Las PyMES en Argentina representan casi el 80% de la industria del software y dada la necesidad de ser competitivas deben mejorar sus métodos y procesos de trabajo. En la mayoría de las empresas el proceso de despliegue no se realiza de manera sistemática y controlada, esto impacta en la finalización del proyecto y la no aceptación del producto final, ocasionando inconvenientes que generan rehacer el trabajo y baja de productividad en su proceso. Ante estas dificultades, se propone realizar un modelo de proceso de despliegue de sistemas de software que permita a las PyMES mejorar la ejecución del proceso de despliegue [4, 18]. En este contexto, la comunicación y la sincronización entre los equipos de trabajo continúan siendo un pilar fundamental para el éxito de un proyecto. La utilización de repositorios de información y sincronización permiten realizar un control de versiones distribuido, trabajando en modo offline o en modo online, con la facilidad de disponer herramientas específicas para la resolución de conflictos entre versiones [3, 4].

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

- Estudio y análisis de procesos de diseño para Bases de Datos no relacionales [1, 2].
- Aplicación y adaptación de procesos de diseño de Bases de Datos no relacionales en casos de usos particulares [28, 33].
- Metodologías y técnicas de la Ingeniería de Software y su aplicación en el desarrollo de Software para escenarios híbridos [2, 34].
- Desarrollo de casos de estudio y métricas de performance para distintos tipos de motores de Bases de Datos No Relacionales (NoSQL) orientados a Familia de Columnas y orientados a Grafos, como, por ejemplo: Apache Cassandra, Neo4j, NebulaGraph y Memgraph, entre otros [14].
- Investigar, desarrollar casos de estudio e incorporar nuevas pruebas para la

escalabilidad horizontal de motores de Bases de Datos No Relacionales [9].
- Repositorios para la sincronización del trabajo en equipo [19, 34].
- Estudio y análisis sobre conceptos básicos de Computación Cuántica.
- Investigar y desarrollar pruebas sobre lenguajes de programación para Computación Cuántica.
- Investigar y desarrollar pruebas sobre simuladores en línea para Computación Cuántica.

3. RESULTADOS OBTENIDOS/ESPERADOS

Los resultados esperados/obtenidos se pueden resumir en:

- Capacitación continua de los miembros de las líneas de investigación.
- Estudio y análisis de procesos de diseño para almacenamiento no estructurado de datos y para Bases de Datos no relacionales [1, 28].
- Análisis comparativo de aplicación de procesos de diseños de Bases de Datos NoSQL. Estudio experimental [31, 33]
- Estudio y análisis de distintos motores de Bases de Datos para el desarrollo de aplicaciones móviles [10, 15, 21].
- Estudio y análisis de Bases de Datos como servicio en la nube [29, 32].
- Análisis sobre la ejecución de consultas en diversos casos de estudio para distintos tipos de motores de Bases de Datos no relacionales [9, 14].
- Estudio y análisis sobre escalamiento horizontal para pruebas de performance entre distintos motores de Bases de Datos no relacionales, principalmente orientado a Grafos [13, 14, 28, 32].
- Análisis de metodologías para la interoperabilidad de sistemas web y aplicaciones móviles [11, 12, 20].
- Análisis comparativo sobre herramientas para programación cuántica. [35]

Algunas de las transferencias realizadas por el III-LIDI relacionadas con este proyecto, entre otras, son:

- Mantenimiento de la aplicación móvil para la comunidad de la Facultad de Informática de la UNLP.
- Mantenimiento del Sistemas de congresos, utilizado por la RedUNCI y otras entidades.
- Mantenimiento de Sistemas de Gestión Administrativa de Instituciones Universitarias.

4. FORMACIÓN DE RECURSOS HUMANOS

Los integrantes de esta línea de investigación dirigen Tesinas de Grado y Tesis de Postgrado en la Facultad de Informática, y Becarios III-LIDI en temas relacionados con el proyecto. Además, participan en el dictado de asignaturas/cursos de grado y postgrado de la Facultad de Informática de la UNLP.

5. BIBLIOGRAFÍA

1. Un estudio de procesos de diseño de bases de datos NoSQL. Marrero Luciano, Osowy Verena, Tesone Fernando, Thomas Pablo Javier, Corbalán Leonardo, Fernández Sosa Juan, Pesado Patricia. XXVIII Congreso

Argentino de Ciencias de la Computación (CACIC 2022) (La Rioja, 3 al 6 de octubre). ISBN: 978-987-1364-31-2. Páginas: 404-414. http://sedici.unlp.edu.ar/handle/10915/149452

2. Aspectos de ingeniería de software, bases de datos relacionales, no relacionales y como servicios en la nube para el desarrollo de sistemas de software híbridos. Marrero Luciano, Thomas Pablo, Pasini Ariel, Bertone Rodolfo, Ibañez Eduardo, Aguirre Verónica, Panizzi Marisa, Olsowy Verena, Tesone Fernando, Pesado Patricia. XXIV Workshop de Investigadores en Ciencias de la Computación (WICC 2022) 28 y 29 de abril, Mendoza. ISBN: 978-987-48222-3-9. Páginas: 370-376. http://sedici.unlp.edu.ar/handle/10915/144325
3. Ingeniería de Software. Un Enfoque Práctico. Novena Edición. Roger S. Pressman, Bruce R. Maxim. McGraw-Hill Interamericana Editores, S.A. 2021. ISBN: 978-1-4562-8772-6.
4. Ingeniería de Software Teoría y Práctica. Pfleegger Shari Lawrence. Pearson / Prentice Hall. 2002. ISBN: 9789879460719
5. Ingeniería de Software. Novena Edición. Ian Sommerville. Addison Wesley / Pearson. 2011. ISBN: 978-607-32-0603-7
6. Diseño Conceptual de Bases de Datos, un enfoque de entidades-interrelaciones. Carlo Batini, Stefano Ceri, Shamkant B. Navathe. Addison-Wesley/Díaz de Santos. ISBN 0-201-60120-6 (1994).
7. Model Driven Extraction of NoSQL Databases Schema: Case of MongoDB. Brahim, A.; Ferhat, R. and Zurfluh, G. (2019). In Proceedings of the 11th International Joint Conference on Knowledge Discovery. V.1: KDIR, ISBN 978-989-758-382-7, pages 145-154.
8. *NoSQL*: modelos de datos y sistemas de gestión de bases de datos. Migani Silvina, Vera Cristian, Lund María Inés. XX Workshop de Investigadores en Ciencias de la Computación (WICC 2018). 26 y 27 de abril, Universidad Nacional del Nordeste (UNNE), Corrientes. ISBN: 978-987-3619-27-4. Páginas: 225-228. http://sedici.unlp.edu.ar/handle/10915/67258
9. Análisis de performance en Bases de Datos NoSQL y Bases de Datos Relacionales. Marrero Luciano, Olsowy Verena, Tesone Fernando, Thomas Pablo, Delia Lisandro, Pesado Patricia. XXVI Congreso Argentino de Ciencias de la Computación (CACIC 2020). ISBN: 978-987-4417-90-9. Páginas: 299-309. Universidad Nacional de La Matanza, del 5 al 9 de octubre del 2020. http://sedici.unlp.edu.ar/handle/10915/114202
10. Un Análisis Experimental de Sistemas de Gestión de Bases de Datos para Dispositivos Móviles. 2021. Tesone Fernando, Thomas Pablo, Marrero Luciano, Olsowy Verena, Pesado Patricia. ISBN: 978-987-633-574-4. Páginas: 343-355. XXVII Congreso Argentino de Ciencias de la Computación (CACIC) (Modalidad virtual, 4 al 8 de octubre de 2021). Universidad Nacional de Salta (UNSA), Argentina. http://sedici.unlp.edu.ar/handle/10915/130353

11. Aspectos de ingeniería de software, bases de datos relacionales y bases de datos no relacionales para el desarrollo de sistemas de software en escenarios híbridos. Luciano Marrero, Pablo Thomas, Ariel Pasini, Rodolfo Bertone, Eduardo Ibáñez, Verónica Aguirre, Verena Olsowy, Fernando Tesone, Patricia Pesado. ISBN: 978-987-3714-82-5. Páginas: 494-498. XXII Workshop de Investigadores en Ciencias de la Computación (WICC 2020). Universidad Nacional de la Patagonia Austral (UNPA). El Calafate, Santa Cruz (Mayo 2020). http://sedici.unlp.edu.ar/handle/10915/104026.
12. Aspectos de ingeniería de software, bases de datos relacionales, y bases de datos no relacionales y bases de datos como servicios en la nube para el desarrollo de software híbrido. Marrero Luciano, Thomas Pablo, Olsowy Verena, Pasini Ariel, Bertone Rodolfo, Ibañez Eduardo, Aguirre Verónica, Panizzi Marisa, Pesado Patricia. XXIII Workshop de Investigadores en Ciencias de la Computación (WICC 2021). ISBN: 978-987-24611-3-3; 978-987-24611-4-0. Páginas: 404-409. Chilecito, La Rioja. http://sedici.unlp.edu.ar/handle/10915/120139
13. Aspectos de ingeniería de software y bases de datos para el desarrollo de sistemas de software en escenarios híbridos. Patricia Pesado, Rodolfo Bertone, Pablo Thomas, Ariel Pasini, Luciano Marrero, Eduardo Ibáñez, Alejandra Rípodas, Verónica Aguirre, Verena Olsowy, Fernando Tesone. XXI Workshop de Investigadores en Ciencias de la Computación (WICC 2019). ISBN: 978-987-3984-85-3. Universidad Nacional de San Juan (UNSJ). Abril 2019. http://sedici.unlp.edu.ar/handle/10915/77088
14. Un estudio comparativo de bases de datos relacionales y bases de datos NoSQL. Pesado Patricia, Thomas Pablo, Delía Lisandro, Marrero Luciano, Olsowy Verena, Tesone Fernando, Fernández Juan Sosa. XXV Congreso Argentino de Ciencias de la Computación (CACIC 2019). Universidad Nacional de Río Cuarto, Córdoba, 14 al 18 de octubre de 2019. ISBN 978-987-688-377-1. http://sedici.unlp.edu.ar/handle/10915/91403
15. A. Nori, "Mobile and embedded databases," in Proceedings of the 2007 ACM SIG- MOD International Conference on Management of data, pp. 1175-1177, 2007.
16. SQLite. https://www.sqlite.org/. Accedido en febrero de 2024.
17. Couchbase. https://www.couchbase.com/. Accedido en febrero de 2024.
18. Aplicaciones para Dispositivos Móviles. Estrategias y enfoques de desarrollo. Thomas Pablo Javier, et. al. XXI Workshop de Investigadores en Ciencias de la Computación (San Juan 2019). ISBN: 978-987-3984-85-3
19. Which Change Sets in Git Repositories Are Related?. Ramadani, J., Wagner, S. EEE 2016. International Conference on Software Quality, Reliability and Security (Viena, Austria.)
20. Synchronization and replication in the context of mobile applications. STAGE, A. (2005 Joint Advanced Student School Course 6: Next-Generation User-Centered).

21. S. Lee, “Creating and using databases for android applications,” International loumal of Database Theory and Application, vol. 5, no. 2, 2012.
22. III-LIDI: https://weblidi.info.unlp.edu.ar/proyectos/investigacion/
23. Firebase. https://firebase.google.com/?hl=es. Accedido en febrero de 2024.
24. Redis. https://redis.io/. Accedido en febrero de 2024.
25. MongoDB. https://www.mongodb.com/es. Accedido en febrero de 2024.
26. Apache Cassandra. https://cassandra.apache.org/_/index.html. Accedido en febrero de 2024.
27. Neo4j. https://neo4j.com/. Accedido en febrero de 2024.
28. Aplicación de Procesos de Diseño de Bases de Datos NoSQL. Marrero Luciano, Olsowy Verena, Tesone Fernando, Thomas Pablo, Pesado Patricia. XXIX Congreso Argentino de Ciencias de la Computación (CACIC 2023). ISBN: 978-987-9285-51-0. Páginas: 134-145 Universidad Nacional de Luján, Luján, Buenos Aires. https://sedici.unlp.edu.ar/handle/10915/163107
29. Performance Analysis in NoSQL Databases, Relational Databases and NoSQL Databases as a Service in the Cloud. Marrero Luciano, Olsowy Verena, Tesone Fernando, Thomas Pablo, Delia Lisandro, Pesado Patricia. https://link.springer.com/chapter/10.1007/978-3-030-75836-3_11
30. Aspectos de ingeniería de Google blog. Presentación del chip cuántico Willow. https://blog.google/technology/research/google-willow-quantum-chip/. Accedido en febrero de 2025.
31. NoSQL database design processes: a comparative study. Marrero Luciano, Olsowy Verena, Thomas Pablo. XI Jornadas de Cloud Computing, Big Data & Emerging Topics. ISBN: 978-950-34-2271-7. Páginas: 58-64. 27 al 29 de junio del 2023, La Plata. http://sedici.unlp.edu.ar/handle/10915/155446
32. Un estudio comparativo de Bases de Datos NoSQL. Luciano Marrero. https://sedici.unlp.edu.ar/handle/10915/155904
33. Un estudio de procesos de diseño de bases de datos NoSQL. Verena Olsowy. https://sedici.unlp.edu.ar/handle/10915/159625
34. Aspectos de ingeniería de software, bases de datos relacionales y bases de datos NoSQL para el desarrollo de sistemas de software híbridos. Marrero Luciano, Thomas Pablo, Pasini Ariel, Bertone Rodolfo, Ibañez Eduardo, Olsowy Verena, Tesone Fernando, Pesado Patricia. XXVI Workshop de Investigadores de la Ciencia de la Computación Puerto Madryn, Chubut, 2024. https://wicc2024.unp.edu.ar/images/libros/LibroDeActasWICC2024.pdf
35. Una revisión de herramientas para programación cuántica. Marrero, Luciano; Olsowy, Verena; Fernández Sosa, Juan; Tesone, Fernando; Corbalán, Leonardo; Thomas, Pablo Javier; Pesado, Patricia. XXX Congreso Argentino de Ciencias de la Computación (CACIC). La Plata, 7 al 11 de octubre de 2024. ISBN: 978-950-34-2428-5. Páginas: 802-813.

http://sedici.unlp.edu.ar/handle/10915/176617
36. Un estudio comparativo de Bases de Datos NoSQL Orientadas a Grafos. Martinez Manuel, Mozzón Federico, Marrero Luciano, Olsowy Verena, Thomas Pablo. XXX Congreso Argentino de Ciencias de la Computación (CACIC). La Plata, 7 al 11 de octubre de 2024. ISBN: 978-950-34-2428-5. Páginas: 591-602. http://sedici.unlp.edu.ar/handle/10915/176463
37. Google Quantum AI. (s.f.). Cirq. Accedido en junio de 2024, de https://quantumai.google/cirq
38. IBM Qiskit SDK Documentation. Accedido en febrero de 2025. https://docs.quantum.ibm.com/guides
39. Microsoft. (s.f.). Introducción a los katas en Q# y QDK. Accedido en junio de 2024, de https://learn.microsoft.com/es-es/azure/quantum/tutorial-qdk-intro-to-katas
40. Amazon Bracket. Accedido en junio de 2024, https://aws.amazon.com/es/braket

Marco Tecnológico para el Desarrollo de Nano Satélites

Fernando Asteasuain – Fernando Pazos- Federico D'Angiolo- Manuel Dubinsky- Matías Loiuseau –Adriana Galli – Martin Miguel Machuca- Ana Jimena Legaspi – Trinidad Mamud - Gastón Freire - Leonardo Guanco
Contacto Principal: fasteasuain@undav.edu.ar
Emails: <InicialNombre.Apellido>@undav.edu.ar

Carrera de Ingeniería en Informática – Universidad Nacional de Avellaneda

Resumen

La industria aeroespacial se ha convertido en los últimos años en una de las áreas de mayor potencial dentro del mundo software tecnológico [1-5]. En particular, el diseño y desarrollo de los denominados nano satélites ha crecido exponencialmente en los últimos dos años. Este tipo de satélites, con bajos costos de desarrollo y de producción, permiten realizar valiosas tareas de producción y recolección de datos en diversos dominios como agroindustria, estudios climáticos, biología especializada o estudios espaciales y astronómicos [1].

El desarrollo de nano satélites constituye un prisma de desafíos tecnológicos y de desarrollo de software que incluye aristas como la verificación formal de software, la exploración de Sistema Auto Adaptativos, la aplicación de algoritmia de Ciencia de Datos, el diseño de vehículos autónomos, la especificación de misiones satelitales, control y comunicación entre dispositivos IoT, procesamiento digital de imágenes a través de mallas poligonales y el manejo del concepto de incertidumbre en el desarrollo de software.

El presente proyecto está enfocado en el desarrollo de una variada gama de herramientas, técnicas y metodologías para encarar los mencionados desafíos.

Palabras claves: *Nano satélites, Computadoras de a bordo, software satelital.*

Contexto

La presente investigación se encuentra enmarcada dentro del proyecto "Marco Tecnológico para el Desarrollo de Nano Satélites" desarrollado en la carrera de Ingeniería en Informática de la Universidad Nacional de Avellaneda.

El grupo de investigación está constituido por tres investigadores formados, cuatro en formación, y cuatro estudiantes de grado haciendo sus tesis de final de carrera y prácticas profesionales. El proyecto entra en 2025 en el tercer año de su desarrollo.

El estudio de OBC para nano satélites ha sido declarado dominio de relevante interés estratégico comprendido en el Plan Nacional de Ciencia, Tecnología e Innovación Argentina 2030.

El proyecto forma parte de un marco inter institucional consistente en poner en órbita un satélite argentino del tipo *CubeSats* [6-9], con el objetivo de promover un laboratorio espacial destinado a la provisión de servicios de *Internet de las Cosas (IoT, por Internet of Things en inglés)*. El grupo de investigación está trabajando conjuntamente con el Consejo Profesional de Ingeniería en Telecomunicaciones, Electrónica y Computación. Finalmente, es importante destacar que existe también la posibilidad de transferencia y servicios al sector industrial y productivo.

1. Introducción

La industria aeroespacial se ha convertido en los últimos años en una de las áreas de mayor potencial dentro del mundo software tecnológico [1-5]. Concretamente, el diseño y desarrollo de los denominados nano satélites ha crecido exponencialmente en los últimos dos años. Este tipo de satélites, con bajos costos de desarrollo y de producción, permiten realizar valiosas tareas de producción y recolección de datos en diversos dominios como agroindustria, estudios climáticos, biología especializada o estudios espaciales y astronómicos [1]. En particular para nuestro país, la investigación y desarrollo de nanosatélites ha sido declarado dominio de relevante interés estratégico comprendido en el Plan Nacional de Ciencia, Tecnología e Innovación Argentina 2030.

El concepto de incertidumbre ha ganado presencia en el desarrollo moderno de software[35-36], y está más que presente en el marco de desarrollar nano satélites. Un claro ejemplo se da en los denominados sistemas auto-adaptativos (Self-Adaptative Sytems o directamente SAS en inglés). Este tipo de sistemas debe no sólo detectar cambios en el entorno o ambiente donde el sistema se está ejecutando, sino también poder cambiar sus objetivos y metas de manera dinámica de acuerdo a esos cambios. Los modelos probabilísticos han demostrado ser una técnica efectiva para incluir el concepto de incertidumbre en el desarrollo de software. Explorar esta técnica dentro del desarrollo satelital constituye un enfoque atractivo a investigar.

En este contexto, el análisis del problema de transmisión de datos desde y hacia el satélite puede ser visto como un problema de control óptimo [31-34]. La dinámica del sistema puede ser identificada a través de datos de entrada/salida oportunamente medidos.

Algunas de las áreas de investigación involucradas son: La verificación formal de software, la Inteligencia Artificial, el procesamiento digital de geometría y la teoría de grafos, diseño de controladores óptimos para la resolución de problemas de optimización condicionada.

.

2. Líneas de Investigación y Desarrollo

En el presente trabajo se explorarán las siguientes líneas de investigación:

- Aplicación de métodos formales de Ingeniería de Software al diseño y verificación de computadoras de abordo para nanosatélites, para artefactos de Internet de las Cosas, protocolos de comunicación, y validación de análisis de ciencia de datos.
- Incluir modelos probabilísticos para poder manejar el concepto de incertidumbre presente en el desarrollo de nano-satélites.
- Facilitar la utilización de patrones de especificación de comportamiento para enunciar formalmente los requerimientos de las misiones satelitales.
- Continuar y profundizar el desarrollo de herramientas de software que den aplicabilidad a los conocimientos adquiridos.
- Analizar y adaptar los nuevos algoritmos de resolución de sistemas laplacianos [7] al contexto de deformación de mallas poligonales.

- Analizar y adaptar los algoritmos aproximados existentes para el problema Low Stretch Spanning Tree [8-10] (LSST) al problema Minimum Spanning Tree Cycle Intersection.

- Adaptar los cuatro algoritmos heurísticos planteados en el contexto del problema Minimum Fundamental Cycle Basis [11] (MFCB) al problema Minimum Spanning Tree Cycle Intersection (MSTCI).
- Analizar los problemas relacionados con la conducción autónoma y su

vínculo con la tecnología V2V y V2I (comunicación entre los vehículos y la infraestructura) [12]. Para tal fin se deben estudiar los desafíos técnicos relacionados con la percepción, la toma de decisiones y la seguridad, que deben resolverse para garantizar que los vehículos autónomos sean seguros y confiables.

- El análisis del problema de transmisión de datos desde y hacia el satélite como un problema de control óptimo. El control óptimo se plantea con el objetivo de minimizar pérdida de datos y energía consumida condicionada a la dinámica de la transmisión, entre otras restricciones que puedan considerarse.
- Estudiar la aplicación de controladores óptimos basados en datos [31-34] objetivando la optimización en la navegación del satélite y/o en la comunicación con el terminal de tierra.

3. Resultados Obtenidos/Esperados

El objetivo principal del presente proyecto de Investigación es el diseño y desarrollo de nano satélites, en especial vinculados al desarrollo de software.

El grupo viene investigando intensamente en los temas involucrados. Por un lado, se ha desarollado un framework de verificación formal y especificación de comportamiento denominado FVS (Feather Weight Visual Scenarios) [15,16]. El mismo cuenta con un lenguaje declarativo, flexible y de gran poder expresivo y ha sido utilizado para validar y sintetizar comportamiento de protocolos de Internet de las Cosas, Planificación y Adaptación de Objetivos para Vehículos Autónomos, Validación de Resultados de Aprendizaje Automático a través de Relaciones Metamórficas, y Síntesis de Comportamiento [14,17,18]. Se ha explorado la utilización de modelos probalísticos dentro de FVS para su utilización en el corto plazo en la especificación de misiones satelitales [30].

En el contexto del problema de deformación de mallas poligonales, implementamos un software basado en los aportes de Matsui [19]. El software permitió realizar una serie de experimentos que derivaron en la definición de un problema novedoso en teoría de grafos: el problema Minimum Spanning Tree Cycle Intersection [19]. En [20-21] avanzamos en esta línea de investigación.

Tomando como la aplicación de Ciencia de Datos en IA el grupo viene estudiando la aplicación de estos algoritmos a sistemas en tiempo real, como por ejemplo, distribución de variables climáticas dentro de Data Centers. Aquí, las variables tales como temperatura, humedad y presión, se cuantifican a través de sensores y luego, por medio de algoritmos basados en Machine Learning, Deep Learning, Control y Optimización, se pueden realizar estudios sobre la distribución y evolución de dichas variables [23-28].

También se estudió el desempeño de diferentes servidores de Internet instalados en la nube que gerencian comunicaciones entre máquinas usando el protocolo *MQTT* [29].

Como resultados futuros esperados mencionar:

a) Estudio de la formulación del problema de comunicación y envío de datos como un problema de control óptimo basado en datos.

b) Exploración de Métodos y Técnicas de Verificación Formal para Ciencia de Datos, y el estudio de Verificación Formal para Protocolos de IoT.

c) Utilización y desarrollo de patrones de especificación de comportamiento para el modelado de requerimientos de misiones satelitales.

d) Implementación de redes Lora para comunicación satelital.

e) Adaptación de métodos heurísticos tomados del problema MFCB para el problema MSTCI y adaptación de métodos aproximados tomados del

problema LSST para el problema MSTCI.

f) Exploración de Métodos y Técnicas algoritmos de conducción autónoma para la toma de decisiones de vehículos y de Métodos que utilizan tecnologías V2V y V2I. Analizar cómo influyen distintos algoritmos, sobre todo de IA.

g) Divulgar los resultados de la investigación en congresos y revistas científicas de interés para los temas de la investigación y analizar y explorar posibilidades de transferencia

h) Consolidar los recursos humanos en inicios de las tareas de investigación como estudiantes avanzados de la carrera, mediante la participación activa en el proyecto o mediante la realización de tesis de final de carrera de grado. También la dirección de tesis de final de carera y supervisión de prácticas profesionales (PPS).

4. Formación de Recursos Humanos

El presente proyecto buscará potenciar a los investigadores en formación y de posgrado.

En cuanto a los estudiantes de grado, se buscará en primer lugar que finalicen su tesis de grado o su práctica profesional. Se los pretende motivar como para que puedan continuar dentro del proyecto mediante la realización de algún posgrado.

Finalmente, se pretende consolidar las líneas de investigación dentro de la carrera de Ingeniería en Informática de la Universidad Nacional de Avellaneda.

5. Bibliografía

1. Saeed, N., Elzanaty, A., Almorad, H., Dahrouj, H., AlNaffouri, T. Y., & Alouini, M. S. (2020). Cubesat communications: Recent advances and future challenges.IEEE Surveys & Tutorials, 22(3), 1839-1862.

2. Hanafi, A., Derouich, A., Karim, M., & Lemmassi, A. (2021, January). Design and Implementation of an Open Source and Low-Cost Nanosatellite Platform. ICDATA (pp.421-432). Springer, Cham.

3. Williams, C. and DelPozzo S. Nano Microsatellite Market Forecast - 10th Edition, Space-Works Annual Nano/Microsatellite Market Assessment, 2020.

4. Pressman, A. (2019). Why Facebook, SpaceX and dozens of others are battling over Internet access from space. Fortune.

5. Kulu E.. "Nanosatellite and cubesat database," 2019. [Online]. Available: https://www.nanosats.eu/

6. Manuel Dubinsky, César Massri y Fernando Asteasuain, "Una metaheurística GRASP para integración en grafos", JAIIO 2017.

7. Spielman D. A. "Algorithms, graph theory, and linear equations in laplacian matrices".

International Congress of Mathematicians, vol. 4, pages 2698–2722, 2010.

8. N. Alon, R. Karp, David Peleg, and Douglas West. A graph-theoretic game and its application to the k-server problem. SIAM Journal on Computing, 24(1):78–100,1995.

9. Ittai Abraham, Yair Bartal, and Ofer Neiman. "Nearly tight low stretch spanning trees". IEEE SFCE; 781–790, 2008.

10. Michael Elkin, Yuval Emek, Daniel A. Spielman, and Shang-Hua Teng, "Lower-stretch spanning trees". ACM, 2005.

11. Narsingh Deo, G. M. Prabhu and M. S. Krishnamoorthy. "Algorithms for generating fundamental cycles in a graph". ACM Trans. Math. Software, 8(1):26–42, 1982.

12. F.M. Tariq , David Isele , JS Baras and Sangjae Bae. SLAS: Speed and Lane Advisory System for Highway Navigation. arXiv:2303.00861v1 [cs.RO] 1 Mar 2023.

13. E. Kaszkurewicz and A. Bhaya. Business Dynamics Models: Optimization-based One Step Ahead Optimal Control. SIAM, 2022

14 Asteasuain, F. (2023). New Horizons for Metamorphic Relationships in Formal

Verification. Communications in Computer and Information Science, vol 1778. Springer, Cham. https://doi.org/10.1007/978-3-031-34147-2_8 .

15 Asteasuain, F., Braberman, V.: Declaratively building behavior by means of scenario clauses. RE 22(2), 239-274 (2017), doi:10.1007/s00766-015-0242-2

16 Asteasuain, F., Calonge, F., Dubinsky, M., Gamboa, P.: Open and branching behavioral synthesis with scenario clauses. CLEI E-JOURNAL 24(3) (2021).

17 F. Asteasuain (2023). "Addressing the State Explosion Problem for Big Data Systems Formal Verication". CACIC 2023

18. F Asteasuain (2023). "Making Self-Adaptive Systems More Verifiable". CONAIISI 2023.

19. Tomomi Matsui, "Algorithms for finding all the spanning trees in undirected graphs", University of Tokyo, 1993.

20. Manuel Dubinsky, César Massri y Gabriel Taubin. "Minimum spanning tree cycle intersection problem". Discrete Applied Mathematics, 294:152–166, 2021.

21. Manuel Dubinsky, César Massri y Gabriel Taubin. "Three aspects of the MSTCI problem", 2023.

22 S. Bae, David Isele, Kikuo Fujimura, and Scott J. Moura.Risk-Aware Lane Selection on Highway with Dynamic Obstacles. arXiv:2104.04105v1 [cs.RO] 8 Apr 2021.

23. Grishina, A., Chinnici, M., Kor, A., Rondeau, E., Georges, J.: A Machine Learning Solution for Data Center Thermal Characteristics Analysis. MDPI-Energies (2020)

24. S. Morteza, M. Nejad., G.Badawy, D.. Down.: Holistic thermal- aware workload management and infrastructure control for heterogeneous data centers using machine learning. FGCS 118 (2021) 208–218 (2021).

25. D'Angiolo, F., Kwist, I., Loiseau, M., Contreras, D., Asteasuain, F.: Algoritmos de Regresión Lineal aplicados al mantenimiento de un Data Center. CACIC 2019

26. D'Angiolo, F., Mas, I., Giribet, J.: Estimación de Temperatura en Servidores mediante Herramientas de Deep Learning. CACIC 2021.

27 D'Angiolo, F., Kwist, I., Mas, I., Giribet, J.: Análisis de Distribución de Temperatura, Humedad y Presión en un Datacenter basado en un Método de Selección de Clusters. CONAIISI 2020.

28 D'Angiolo, F., Kwist, I., Loiseau, M., Contreras, D., Glas, G.: Algoritmo de KNN Aplicado al Mantenimiento de un Datacenter. CONAIISI 2019.

29 Pazos, F. Performance evaluation of MQTT broker servers deployed in the cloud. JAIIO (2024, en prensa).

30. Asteasuain F, Legaspi A, Freire G, Machuca M "A Declarative and Probabilistic Behavioral Specification Language to Handle Uncertainty in Modern Software Development" CONAIISI 2024, Catamarca.

31. Vangelista, Lorenzo: "Frequency shift chirp modulation: The Lora modulation". IEEE Signal Processing Letters, volume 24, number 12, pages: 1818-1821, 2017

32. De Persis, C, and Tesi, P. "Formulas for Data-Driven control:Stabilization, Optimality and Robustness". IEEE Transactions on Automatic Control, vol. 65, n. 3, 2020. DOI: 10.1109/TAC.2019.2959924

33 Dorfler, F., Tesi, P. and De Persis, C. "On the Certainty-Equivalence Approach to Direct Data-Driven LQR Design. IEEE Transaction on Automatic COntrol, vol.68,n. 12, 2023. DOI: 10.1109/TAC.2023.3253787

34. Zhao, F., Dorfler, F. and You, K. "Data-enabled Policy Optimization for the Linear Quadratic Regulator". (CDC),pp 6160-6165, December 2023, Singapore. DOI: 10.1109/CDC49753023.1038347.0

35. Vogel, T., Carwehl, M., Rodrigues, G. N., & Grunske, L. (2023). A property specification pattern catalog for real-time system verification with UPPAAL. Information and Software Technology, 154, 107100.

[36] Zudaire, S., Gorostiaga, F., Sánchez, C., Schneider, G., & Uchitel, S. (2021, May). Assumption monitoring using runtime verification for UAV temporal task plan executions. In (ICRA) (pp. 6824-6830). IEEE.

RECUPERACIÓN Y REUSO DE INFORMACIÓN RELEVANTE EN COMUNIDADES VIRTUALES

Nadina Martinez Carod[1,2], Gabriela Aranda[1,2], Valeria Zoratto[1,2],
Natalia Baeza[1], Lucas Cavaliere[1], Sandra Lucero[1], Nahuel Mamani[1]

[1] Facultad de Informática. Universidad Nacional del Comahue
Buenos Aires 1400, (8300) Neuquén
[2] GIISCo, Grupo de Investigación en Ingeniería de Software del Comahue
{nadina.martinez, gabriela.aranda,vzoratto}@fi.uncoma.edu.ar

RESUMEN

En la era digital actual, el intercambio de experiencias entre usuarios de la web ha alcanzado un nivel sin precedentes. A través de plataformas tales como las redes sociales, blogs, wikis y otras comunidades en línea, se fomenta al máximo la colaboración, permitiendo compartir opiniones, experiencias y contenido de manera dinámica y participativa, donde todos los usuarios contribuyen activamente en la creación y socialización de contenido.

La evolución hacia la Web 3.0, o Web semántica, busca expandir la funcionalidad de la Web 2.0 incorporando la interpretación computarizada de la información presente en dicha red. Este enfoque agrega inteligencia mediante técnicas de aprendizaje automático, permitiendo a los usuarios de la web experiencias más personalizadas.

En este contexto, nuestro proyecto se centra en el aprovechamiento de la información proveniente de las comunidades virtuales de preguntas y respuestas (CQA, del inglés *Community Question Answering*) como entornos colaborativos para la adquisición de conocimientos. Específicamente, nos enfocamos en la reutilización de respuestas a preguntas previas, enlaces a recursos externos y la combinación de múltiples fuentes de información, para proporcionar a los usuarios de dichas comunidades virtuales, respuestas más completas y precisas.

Nuestro enfoque no solo implica el uso de técnicas de detección, análisis de información y algoritmos de recomendación, sino también la provisión de herramientas para el reuso y manipulación de la información propia de las comunidades.

Palabras clave: Recuperación de información, información en CQA, comunidades virtuales, reuso de información.

CONTEXTO

Nuestro proyecto de investigación, titulado “Reuso de información en comunidades virtuales” (04/F018), a realizarse en el periodo 2022-2026, surge como continuación de una de las líneas del Programa 04/F001 de la Universidad Nacional del Comahue, “Desarrollo de Software basado en Reuso”, que se llevó a cabo entre 2013 y 2021. Dicha línea de investigación estaba centrada en el reuso de información proveniente de foros de discusión técnicos.

El alcance de nuestro proyecto actual se extiende para abarcar otros tipos de comunidades virtuales, ampliando también el dominio de aplicación a otras áreas de la informática, educación y otras áreas sociales. Nuestro objetivo principal es avanzar, a partir del conocimiento ya adquirido, para aportar a

la toma de decisiones basada en información disponible en la web, recuperada y procesada.

1. INTRODUCCIÓN

La búsqueda de información en la Web ha experimentado una transformación significativa por la creación constante de nuevo contenido mediante la interacción de los usuarios en línea. Este fenómeno, especialmente evidente en entornos digitales y plataformas en línea, ha dado lugar a la emergencia y proliferación de lo que se conoce como Comunidades Virtuales [13]. Dentro de este grupo, las Comunidades de Preguntas y Respuestas (CQA) se destacan como las más ampliamente utilizadas para la recuperación de información.

Desde la llegada de la Web 2.0, diversos estudios han explorado las oportunidades que estas tecnologías ofrecen para la adquisición de conocimiento [1], sin embargo, el crecimiento masivo de información en estos sitios genera desafíos para garantizar la precisión de las respuestas a las consultas planteadas, motivando diversas líneas de investigación [15]. Estas investigaciones abarcan diversos aspectos, como el estudio de los usuarios (características, motivación, participación, interacción y roles) [14], la evaluación de la calidad de las respuestas brindadas sobre las consultas [2, 5, 9], la veracidad de la información contenida en los posts [12], la gestión de la información disponible (como técnicas para evitar duplicación de información y organización de la misma) [18], así como las referidas a la interacción social, es decir a cómo se construyen las comunidades sobre temas específicos [7].

Para analizar estas comunidades se pueden emplear conjuntos de datos recopilados de las CQA, que incluyen mensajes de foros, comentarios en blogs y publicaciones en redes sociales, junto con información de los usuarios. Estos conjuntos de datos pueden ser corpus ya disponibles para investigaciones académicas o bien el producto de la aplicación de herramientas que permiten analizar datos, basados en modelos de lenguaje natural. Por ejemplo, algunos estudios utilizan la API de StackOverflow[1] obtener corpus de datos específicos, que incluyen contenido sobre preguntas, respuestas, usuarios, etiquetas, votos, etc. así como otro tipo de información disponible en los sitios de StackExchange [8, 17].

En cuanto a estudios realizados sobre dichos corpus, se ha analizado la relación a la construcción de conocimiento a partir de los diálogos entre usuarios, investigando las posibles preguntas aclaratorias formuladas por éstos en diferentes dominios [9]. De esta manera se apunta a predecir una pregunta de manera proactiva aclaratoria, cuando la intención de un usuario no está clara, lo que podría ayudar al sistema a proporcionar respuestas más útiles.

Por otro lado, se ha explorado la medición y recuperación de la participación ciudadana a través de estas plataformas de comunidades virtuales con el objetivo de determinar cómo la información recuperada puede ser utilizada para la toma de decisiones de entidades gubernamentales, realizadas a partir de la opinión de la ciudadanía en general. Con este objetivo en mente, es necesario analizar las herramientas existentes (como Decim[2] y Consul[3]) [14, 16], así como diseñar y desarrollar nuevas herramientas, mediante sitios web o aplicaciones móviles específicas, que le den soporte.

Dado que estas tecnologías se basan en recuperar opiniones de ciudadanos, su análisis y evaluación son similares a las comunidades virtuales ya analizadas en este proyecto, por lo

[1] https://api.stackexchange.com/
[2] https://decidim.org/
[3] https://consuldemocracy.org/

que en este proyecto apuntamos a hacer uso de las técnicas y herramientas elaboradas previamente, como soporte para la toma de decisiones basada en la opinión de ciudadanos, aplicados al dominio de los barrios de la ciudad de Neuquén.

En resumen, nuestro proyecto se enfoca en recuperar información disponible en las comunidades virtuales, que puede utilizarse mediante técnicas de reuso e integración de la información (IRI, del inglés *Information Reuse and Integration*) para obtener, integrar y validar información existente hasta la fecha, con el fin de aplicarla en los procesos de toma de decisiones en diferentes dominios [6].

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El proyecto cuenta con la participación de docentes y estudiantes del Grupo de Investigación en Ingeniería de Software del Comahue (GIISCo), perteneciente a la Facultad de Informática de la Universidad Nacional del Comahue (UNCo). Además, se suma la colaboración de asesores externos de la Facultad de Humanidades de la misma universidad, y otra asesora de la Facultad de Ciencias Exactas de la Universidad Nacional del Centro de la Provincia de Buenos Aires (UNCPBA).

El enfoque principal del proyecto radica en definir técnicas y algoritmos de recomendación para proporcionar asistencia inteligente a los usuarios de comunidades virtuales en su búsqueda de información relevante. Los docentes participantes son especialistas en diversos campos, incluyendo Programación, Ingeniería en Sistemas y Teoría de la Computación, lo que enriquece la diversidad de perspectivas y enfoques aplicados en el desarrollo del proyecto.

Esta variedad de disciplinas entre los docentes involucrados permite abordar los objetivos específicos del proyecto desde diferentes criterios, lo que facilita la implementación de mejoras y promueve el trabajo colaborativo entre los miembros del equipo de investigación.

3. RESULTADOS OBTENIDOS Y ESPERADOS

Nuestras líneas de investigación, como parte del Grupo de Investigación en Ingeniería de Software del Comahue (GIISCo), están basadas en tópicos relacionados con la Ingeniería de Software, abordando aspectos aplicados a reuso de conocimiento, desde el año 2013. En dicho momento, se comenzó presentando modelos conceptuales para la clasificación de la información contenida en foros de discusión técnicos, bajo el perfil del usuario que utiliza o consume la información, en el rol de usuario externo de dichas comunidades virtuales [3].

A continuación se trabajó sobre los foros de discusión como un tipo específico de las comunidades virtuales, no sólo en referencia a la clasificación de los roles de usuarios de acuerdo al conocimiento y expertitud de los mismos [11], sino también para la propuesta de métricas de calidad para el contenido de los hilos de discusión [10].

También se aplicaron diversas herramientas gramaticales para la clasificación del contenido de foros de discusión [4], sumado a la inclusión de una base de datos léxica y un analizador morfológico para la detección de palabras de acuerdo a su uso en el contexto. Asimismo, se analizaron hilos de discusión técnicos para encontrar un conjunto mínimo de características que puedan utilizarse para clasificar un post según su calidad, teniendo en cuenta diferentes clasificadores para no sesgar las pruebas y resultados [19].

A partir de lo obtenido en trabajos previos, se ha avanzado en la creación de nuevos modelos de aprendizaje automático para distintos

aspectos de la información disponible en foros de discusión, como clasificación de mensajes [21] y clasificación de usuarios [22], y se continúa ampliando el conjunto de técnicas y herramientas utilizadas. Por otro lado, en colaboración con el ISISTAN, se planea seguir en la búsqueda de sistemas de asistencia inteligente para el reuso de conocimiento, y su aplicación para la recomendación de piezas de información en comunidades virtuales.

Por último, como enfoque práctico, estamos trabajando para aplicar los resultados de nuestra investigación en aspectos sociales específicos de nuestra región, lo que nos motiva diseñar propuestas para la toma de decisiones basadas en la recuperación de información sobre opinión ciudadana con herramientas diseñadas específicamente para las comunidades virtuales de los barrios de la ciudad de Neuquén.

4. FORMACIÓN DE RECURSOS HUMANOS

El proyecto cuenta con la colaboración de profesionales, asesores e investigadores del CONICET, docentes de informática, y estudiantes de licenciatura y maestría que están desarrollando sus tesis en el proyecto. Esta diversidad de participantes enriquece el enfoque multidisciplinario del mismo. Entre los miembros destacados de la Universidad Nacional del Comahue se encuentran:

- Tres docentes del Departamento de Programación con dedicación exclusiva y Doctorado en Ciencias de la Computación.
- Una docente del Departamento de Programación con dedicación parcial, realizando su tesis de Maestría en Ciencias de la Computación.
- Un docente del Departamento de Ingeniería de Computadoras con dedicación simple.
- Una docente colaboradora del Departamento de Teoría de la Computación con dedicación exclusiva.

Además, se cuenta con la asesoría de una profesional del Instituto Superior de Ingeniería del Software (ISISTAN) de la Universidad Nacional del Centro de la Provincia de Buenos Aires (UNCPBA), con experiencia en Sistemas de Recuperación de Información y doctorado en Ciencias de la Computación. También colaboran docentes de la Facultad de Humanidades y de la Facultad de Economía y Administración del Instituto Patagónico de Estudios de Humanidades y Ciencias Sociales.

El proyecto también cuenta con la participación de una estudiante que está realizando su tesis de Maestría en Ciencias de la Computación, tres estudiantes que están realizando su tesis de Licenciatura en Ciencias de la Computación, y se planea incorporar a dos más durante el presente año.

5. BIBLIOGRAFÍA

[1] Al-ghamdi, H. A. K., & Al-ghamdi, A. A. K. (2015). The role of virtual communities of practice in knowledge management using Web 2.0. *Procedia Computer Science*, *65*, 406-411.

[2] Amancio L., Dorneles C., and Dalip D.. Recency and quality-based ranking question in CQAs: A stack overflow case study. Information Processing Management, 58(4), 2021.

[3] Aranda, G. N., Martínez Carod, N., Faraci, P., & Cechich, A. (2013). Hacia un framework de evaluación de calidad de información en foros de discusión técnicos. In *XIV Simposio Argentino de Ingeniería de Software (ASSE)-JAIIO 42 (2013)*.

[4] Aranda G, Zoratto V., Martínez Carod N., Roger S, Otermin F., Cechich A. . Clasificación de contenido de hilos de discusión mediante análisis sintáctico y morfológico. CICCSI 2018. ISBN 9789874568366. Mendoza, 2018, pp. 35-44.

[5] Braslavski, P., Savenkov, D., Agichtein, E., & Dubatovka, A. (2017, March). What do you mean exactly? Analyzing clarification questions in CQA. In *Proceedings of the 2017 conference on*

conference human information interaction and retrieval (pp. 345-348).

[6] Day, M. Y., Ong, C. S., & Hsu, W. L. (2010). An analysis of research on information reuse and integration (2003-2008). *International Transactions on Systems Science and Applications*, *6*(2/3), 146-157.

[7] Fu, C. (2020). User intimacy model for question recommendation in community question answering. *Knowledge-Based Systems*, *188*, 104844.

[8] Hazoom, M., Malik, V., & Bogin, B. (2021). Text-to-SQL in the wild: a naturally-occurring dataset based on Stack exchange data. *arXiv preprint arXiv:2106.05006*.

[9] Lu, H., & Kong, M. (2017, February). Community-based question answering via contextual ranking metric network learning. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 31, No. 1).

[10] Martínez Carod, N. M., Zoratto, V., Aranda, G., & Faraci, P. (2018). Aplicación de Métricas de calidad a hilos de discusión. CICCSI 2018. ISBN 9789874568366. Mendoza, 2018, pp. 151-160.

[11] Martínez Carod, N. , Aranda, G., Zoratto, V., & Murray, C.. Una propuesta para clasificación de roles de usuarios en foros de discusión técnicos. CACIC (2019).

[12] Mihaylova, T., Karadjov, G., Atanasova, P., Baly, R., Mohtarami, M., and Nakov, P. (2019). SemEval-2019 task 8: Fact checking in community question answering forums. In Proceedings of the 13th International Workshop on Semantic Evaluation, pages 860–869, Minneapolis, Minnesota, USA. Association for Computational Linguistics.

[13] Ogbamichael, H. B., & Warden, S. (2018). Information and knowledge sharing within virtual communities of practice. South African Journal of Information Management, 20(1), 1-11.

[14] Rivera Rásury, M. X. (2019). Desarrollo de una herramienta de soporte metodológico a los procesos de e-participación.

[15] Roy, P. K., Saumya, S., Singh, J. P., Banerjee, S., & Gutub, A. (2023). Analysis of community question-answering issues via machine learning and deep learning: State-of-the-art review. CAAI Transactions on Intelligence Technology, 8(1), 95-117.

[16] Santamaría-Philco, A., Cerdá, J. H. C., & Gramaje, M. C. P. (2019). Advances in e-Participation: A perspective of Last Years. *Ieee Access*, *7*, 155894-155916.

[17] Wang, S., Lo, D., & Jiang, L. (2013, March). An empirical study on developer interactions in stackoverflow. In *Proceedings of the 28th annual ACM symposium on applied computing* (pp. 1019-1024).

[18] Zhang, Y., Lo, D., Xia, X., & Sun, J. L. (2015). Multi-factor duplicate question detection in stack overflow. *Journal of Computer Science and Technology*, *30*, 981-997.

[19] Zoratto V., Aranda G, Martinez Carod N., Otermin F.. Evaluación de estrategias para clasificar hilos de foros de discusión según su contenido, ASSE-JAIIO 2021, Argentine Symposium on Software Engineering, Argentina, 2021.

[20] Zoratto, V., Godoy, D., & Aranda, G. N. (2023). A Study on Influential Features for Predicting Best Answers in Community Question-Answering Forums. Information, 14(9), 496.

[21] Zoratto, V., Aranda, G. N., Saldivia Urra, J. (2024) Clasificación de posts según su rol en hilos de foros de discusión, CACIC 2024: XXX Congreso Argentino de Ciencias de la Computación, Oct 2024, La Plata, ISBN: 978-950-34-2428-5, pp. 678-687

[22] Aranda, G., Zoratto, V., Cárdena, A. (2024): Análisis del comportamiento de los usuarios que preguntan en CQA, CACIC 2024: XXX Congreso Argentino de Ciencias de la Computación, Oct 2024, Plata, ISBN: 978-950-34-2428-5, pp. 629-638

Reestructuración de Escenarios hacia una Visión Técnica

María C. Elizalde[1], Jorge H. Doorn[1], Graciela D.S. Hadad[1,2]

[1] Instituto de Ingenierías y Nuevas Tecnologías, Universidad Nacional del Oeste
[2] Facultad de Ingeniería y Tecnología Informática, Universidad de Belgrano
{melizalde, jdoorn, ghadad}@uno.edu.ar

RESUMEN

La Ingeniería de Requisitos procura describir los servicios a ser prestados por el futuro sistema de software desde el punto de vista de la interacción de este con los restantes componentes del contexto en el que se desempeñará. Esto involucra: i) enfatizar la interacción entre los seres humanos y el sistema, ii) organizar su descripción en base a los procesos del negocio, y iii) utilizar el vocabulario específico de ese contexto, lo que lleva al uso de modelos escritos en lenguaje natural, tal como el modelo de escenarios. Las ventajas dadas por estas descripciones que reducen los problemas de comunicación entre las personas, originan ciertas dificultades en las actividades posteriores en el proceso de desarrollo de software. Resulta entonces necesario que estas descripciones se acerquen a las descripciones más técnicas, en términos de las características del sistema de software en sí mismo. Se propone elaborar mecanismos de transformación o reestructuración de escenarios para facilitar actividades, tales como asegurar la calidad de los requisitos, planificando y definiendo mecanismos de validación, priorizar los servicios a ser prestados por el sistema, definir la arquitectura del software, o establecer un orden de desarrollo de sistema.

Palabras clave: Ingeniería de Requisitos, Escenarios, Calidad de Modelos, Priorización, Arquitectura del Software.

CONTEXTO

La propuesta que se presenta describe los lineamientos del proyecto de investigación acreditado "Evolución de los requisitos desde una visión organizacional hacia una visión técnica." del Instituto de Ingeniería y Nuevas Tecnologías de la Universidad Nacional del Oeste.

1. INTRODUCCIÓN

En el proceso de requisitos a utilizar en el marco del proyecto de investigación, los servicios a ser prestados por el sistema de software se presentan mediante dos modelos que utilizan narrativas en lenguaje natural: los escenarios futuros y el documento de especificación de requisitos (Leite et al., 2004). El primero es esencial para comprender las propiedades del sistema, mientras que el segundo, opcional, es principalmente de características contractuales. Cada escenario describe una situación en el contexto de uso del sistema, presentado desde el punto de vista de ese contexto, donde se describe muestra el comportamiento de los actores, el sistema de software, y otros sistemas o dispositivos. Cabe notar que los requisitos del software están implícitamente definidos en este conjunto de escenarios (Hadad et al., 2009).

Este proceso de requisitos basado en modelos en lenguaje natural organiza los escenarios en secuencias de cómo se llevarán a cabo los procesos de negocio en el contexto de uso (Hadad, 2008). Utiliza una organización jerárquica, donde los escenarios de alto nivel suelen representar procesos de negocio y los de más bajo nivel actividades y sub-actividades de esos procesos (ver Figura 1).

Sin embargo, la organización de estos escenarios no siempre es la apropiada, dependiendo del uso que se les dará. Así que, para obtener esa visión técnica necesaria, se deberá comenzar por la reorganización o agrupamiento de los escenarios, en función de la actividad en la que van a ser utilizados. Posiblemente, en algunos casos, sea necesario realizar alguna transformación menor para

mejorar la visualización o atemperar las ambigüedades que pudieran surgir de esa reestructuración.

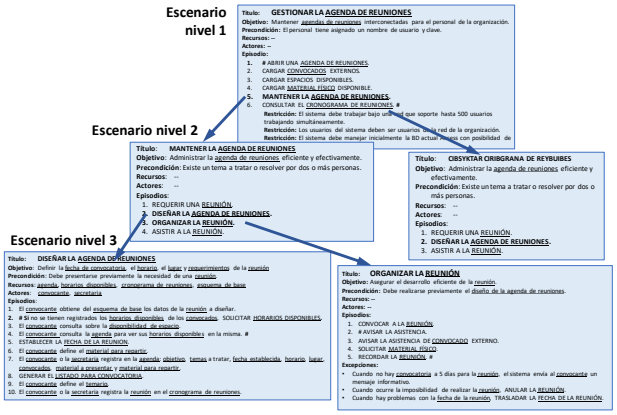


Figura 1. Organización Jerárquica de Escenarios

En el presente proyecto se aspira a definir mecanismos precisos de reestructuración de escenarios futuros para facilitar el desarrollo de actividades que suelen partir de este modelo, tales como asegurar la calidad de los requisitos, planificando y definiendo mecanismos de validación, priorizar servicios a ser prestados por el sistema, definir la arquitectura del software, o establecer un orden de desarrollo de sistema.

Con ello se pretende determinar cuáles son las transformaciones que es apropiado realizar sobre el conjunto de escenarios futuros, tanto para llevar a cabo las actividades reflexivas como para presentarlos en una forma más fácil de interpretar para quien está abocado a la construcción del sistema de software.

Esta reestructuración de escenarios cubrirá una brecha importante, existente entre la Ingeniería de Requisitos y las restantes actividades en el desarrollo de software, escasamente tratada en la literatura (Bastos & Castro, 2003) (Bidian & Yu, 2007) (de Oliveira & Silva, 2015).

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Como ocurre en numerosas disciplinas en las que tienen lugar procesos de producción, los requisitos del producto a elaborar no se corresponden exactamente con los deseos o aspiraciones de los clientes y usuarios, ni con las pretensiones de los productores, sino que resultan ser una amalgama de las aspiraciones de todos los involucrados, configurada por normativas legales, el estado del arte y las reglas de las buenas prácticas de la disciplina. En virtud de la alta flexibilidad inherente, lo anterior es especialmente vigente en el caso de producción de sistemas de software.

Una consecuencia inmediata de lo anterior es que determinar los servicios a ser prestados por un sistema de software no es una actividad indagatoria en la que el ingeniero de requisitos averigua que es lo que hay que hacer, sino que, por el contrario, es una actividad creativa en la que se definen las características del producto de software a producir (Hadad, 2008). Obviamente, quien lleva adelante esta tarea debe conocer las posibilidades de las herramientas que se van a utilizar, pero también debe conocer acabadamente las aspiraciones y necesidades de los clientes y usuarios. Finalmente, de estar estos disponibles, los servicios a ser prestados deben ser acordados con los clientes y usuarios, lo que puede involucrar un proceso de negociación que debe considerar los intereses y las dificultades de todos los involucrados (Hadad, 2008).

No hay acuerdo ni negociación posible si no se aborda el tratamiento del tema utilizando un lenguaje común sin ambigüedades ni sobreentendidos que lleven a malas interpretaciones. Es por esta razón que la Ingeniería de Requisitos ha privilegiado el uso del lenguaje natural en sus presentaciones, incluso utilizando el vocabulario del contexto donde se desempeñará el sistema de software, minimizando los tecnicismos propios de la disciplina (Leite et al., 2004) (Kaplan et al., 2013) (Antonelli et al., 2021).

Profusas experiencias, tanto académicas como aquellas obtenidas en la práctica real, dan muestra de lo acertado de esta estrategia (Mighetti & Hadad, 2016) (Mira et al., 2018). Sin embargo, los modelos escritos en lenguaje natural, en los términos propios del contexto, no son completamente apropiados para llevar a cabo las primeras actividades de construcción de software propiamente dicho.

La alternativa al uso del lenguaje natural ya ha sido probada en numerosas ocasiones y

básicamente consiste en describir los servicios a ser prestados por el sistema de software mediante esquemas gráficos, especificaciones de algoritmos y otros elementos propios de las Ciencias de la Computación (Bidian & Yu, 2007) (Ordinez et al., 2010) (Figueiredo et al., 2024). De esta forma es necesario entrenar a los clientes y usuarios en la interpretación de estas representaciones, o realizar traducciones que les faciliten su comprensión. La experiencia ha mostrado que los requisitos que se definen de esta manera suelen contener una cantidad inaceptable de errores (Jackson, 1995).

El beneficio de mejorar tempranamente la calidad de los requisitos del software es innegable, por lo que es preferible posponer el uso de representaciones que faciliten actividades posteriores, garantizando que toda modificación a los requisitos, ya sea por cambios en el contexto como por cambios tecnológicos, se sigan presentando y evaluando desde el punto de vista del contexto de uso del sistema.

En (Zowghi & Gervasi, 2002) se explora cómo los requisitos expresados en lenguaje natural influencian la arquitectura del software, discutiendo como reelaborar dichos requisitos para un apropiado diseño del software, mientras que en (Liu & Shih, 2003) se propone una forma de trasladar requisitos expresados en lenguaje natural a diagramas UML. En el mismo sentido, en la tesis de maestría de Chen (Chen, 2008) se presenta una herramienta para realizar dicha traducción.

Por su parte, en (Berry et al., 2003) se evalúan varios métodos para transformar requisitos escritos en lenguaje natural en modelos más formales. En (Kop et al., 2010), se propone un mecanismo para traducir requisitos en lenguaje natural a modelos conceptuales. En (Maltempo et al., 2024) se propone un mecanismo similar partiendo de especificaciones en lenguaje natural y utilizando una herramienta basada en procesamiento de lenguaje natural para producir un modelo conceptual. Utilizando un enfoque diferente, en (Kuhn, 2014) se comparan diferentes usos de subconjuntos del lenguaje natural, que introducen restricciones gramaticales y de vocabulario para facilitar la traducción a especificaciones más formales.

Con el propósito de facilitar la verificación de requisitos escritos en lenguaje natural, en (Fantechi et al., 2003) se evalúan herramientas y técnicas lingüísticas para mejorar la consistencia y reducir la cantidad de errores en requisitos.

En el dominio de las aplicaciones de bases de datos, en (Tjoa & Berger, 1994) se describe una estrategia para construir un modelo entidad-relación a partir de requisitos descriptos en lenguaje natural.

Como se puede observar, existe un conjunto importante de contribuciones que destacan la importancia de elaborar, documentar, verificar y validar los requisitos utilizando representaciones basadas en el uso del lenguaje natural, con o sin restricciones gramaticales. Sin embargo, prácticamente todas ellas avanzan en la construcción de representaciones más apropiadas para el desarrollo del sistema de software utilizando como punto de partida los requisitos propiamente dichos. En el presente proyecto se intentará comprobar la hipótesis que es mucho más apropiado partir de conjunto de escenarios futuros, ya que estos expresan los requisitos en su contexto con el importante desambiguado que provee tanto la estructura de esos escenarios como las aclaraciones acerca de todos los términos propios de ese contexto disponibles como parte de ellos.

En otras palabras, se pretende conservar toda la riqueza contextual del conjunto de escenarios, facilitando al mismo tiempo las tareas posteriores, con solo aplicar mecanismos que permitan diferentes formas de presentación de los escenarios futuros.

Asimismo, se explorará la posibilidad de utilizar herramientas de Inteligencia Artificial basadas en procesamiento de lenguaje natural (Dalpiaz & Niu, 2020) (Ahmad et al. 2021) (Kaur et al., 2021), aprovechado la semántica contextual de los escenarios, especialmente aquella basada en el glosario de los términos específicos de ese contexto, asociado a los escenarios.

El objetivo general del presente proyecto es establecer mecanismos sistematizados de reestructuración de escenarios para facilitar actividades de aseguramiento de calidad y de construcción de software.

Se han planteado los siguientes objetivos específicos:

- Identificar las actividades de aseguramiento de calidad del conjunto de escenarios futuros que son afectados por la propia organización del conjunto de escenarios futuros.
- Identificar actividades del proceso de software que utilizan el conjunto de escenarios futuros y qué tipo de información precisan.
- Determinar cuál es la forma más apropiada de presentación de la información almacenada en los escenarios futuros según las actividades identificadas.
- Definir las transformaciones necesarias en el conjunto de escenarios según cada actividad del proceso de software identificada, factorizando aspectos comunes de esas transformaciones.
- Explorar el uso de herramientas de inteligencia artificial basadas en procesamiento de lenguaje natural para una posible automatización de las reestructuraciones.

Las hipótesis de trabajo se basan en: i) la suposición que el conjunto de escenarios es un mejor punto de partida para obtener tanto representaciones que faciliten la validación de requisitos como representaciones con perfiles más técnicos apropiados para las actividades de diseño y construcción del sistema de software, y ii) que es posible realizar una reorganización semánticamente invariante del conjunto de escenarios.

Se asume que la reestructuración de escenarios futuros es una actividad viable, de costo relativo y que puede sistematizarse mediante una serie de transformaciones.

Se considera que ello permitirá mejorar la productividad de las actividades que usan este modelo, tales como la validación temprana de requisitos, su priorización y la posterior implementación de requisitos en componentes de software, así como una mejor organización de las pruebas de aceptación.

3. RESULTADOS OBTENIDOS/ESPERADOS

En la propuesta inicial del presente proyecto se espera obtener al finalizar el mismo, los siguientes resultados:

i. procedimientos para la reestructuración de escenarios futuros dirigidos a actividades de validación de requisitos, priorización, identificación del tipo de arquitectura del software y ordenamiento para el desarrollo del software, y

ii. pautas para el uso de herramientas de inteligencia artificial para colaborar en el agrupamiento de escenarios futuros y su visualización.

4. FORMACIÓN DE RECURSOS HUMANOS

En el proyecto participan tres investigadores, y se prevé la continuidad de dos alumnos de grado, participantes en el proyecto anterior. También se ha incorporado un alumno que desarrollará su Trabajo Final de carrera dentro de los alcances de este proyecto.

5. BIBLIOGRAFÍA

Ahmad, K., Bano, M., Abdelrazek, M., Arora, C., & Grundy, J. (2021) What's up with Requirements Engineering for Artificial Intelligence Systems? 2021 IEEE 29th International Requirements Engineering Conference (RE), Notre Dame, IN, USA, pp. 1-12.

Antonelli, L., Leite, J.C.S.P., Oliveros, A., Rossi, G. (2021). Specification Cases: a lightweight approach based on Natural Language. 24th Workshop on Requirements Engineering (WER21), Brasilia, Brazil.

Bastos L.R.D, & Castro, J.F.B. (2003). Integration between Organizational Requirements and Architecture. 6th Workshop on Requirements Engineering (WER03), Piracicaba, Brasil, pp. 124-139.

Berry, D. M., Kamsties, E., & Krieger, M. (2003). From natural language requirements to rigorous requirements: Bridging the gap with techniques and tools. En Formal Methods for Components and Objects.

Bidian, C., & Yu, E. (2007). Towards Variability Design as Decision Boundary Placement. 10th Workshop on Requirements Engineering (WER07), Toronto, Canadá, pp. 139-148.

Chen, L. (2008). Computer-assisted transformation of design documents from a natural language description to structured modeling languages. Master thesis, Concordia University, Canadá.

Dalpiaz, F., & Niu, N. (2020). Requirements Engineering in the Days of Artificial Intelligence. En IEEE Software, Vol. 37, No. 4, pp. 7-10.

de Oliveira. O.S., & Silva, C. (2015). Modelando SOA a partir de Modelos Organizacionais. 18th Workshop on Requirements Engineering (WER15), Lima, Perú.

Fantechi, A., Gnesi, S., Lami, G., & Maccari, A. (2003). Applications of linguistic techniques for use case analysis. Requirements Engineering Journal, 8(3), 161-170.

Figueiredo, R.C., & Claro, D.B., Maciel, R.S.P., Leite, J.C.S.P. (2024). Using I* For an Early Analysis of Interoperability Requirements. 27th Workshop on Requirements Engineering (WER24), Buenos Aires, Argentina.

Hadad, G.D.S. (2008). Uso de Escenarios en la Derivación de Software. Tesis Doctoral, Universidad Nacional de La Plata.

Hadad, G.D.S., Doorn, J.H., Kaplan, G.N. (2009). Explicitar Requisitos del Software usando Escenarios. 12th Workshop on Requirements Engineering (WER09), Valparaíso, Chile, pp. 63 -74.

Jackson, M. (1995) Software Requirements & Specifications. A lexicon of practice, principles and prejudices. Addison-Wesley, Reading, MA/ACM Press, Nueva York.

Kaplan, G., Doorn, J.H., Gigante, N. (2013). Evolución semántica de glosarios en los procesos de requisitos. XVIII Congreso Argentino de Ciencias de la Computación, Mar del Plata, pp. 777-786.

Kaur, K., Singh, P., Kaur, P. (2021). A Review of Artificial Intelligence Techniques for Requirement Engineering. En: Singh, V., Asari, V.K., Kumar, S., Patel, R.B. (eds) Computational Methods and Data Engineering. Advances in Intelligent Systems and Computing, Vol 1257. Springer, Singapur.

Kop, C., Fliedl, G., & Mayr, H.C. (2010). From natural language requirements to a conceptual model. International Workshop on Design, Evaluation and Refinement of Intelligent Systems (DERIS2010), vol. 67.

Kuhn, T. (2014). A survey and classification of controlled natural languages. Computational Linguistics, 40(1), 121-170.

Leite, J.C.S.P., Doorn, J.H., Kaplan, G.N., Hadad, G.D.S., Ridao, M.N. (2004). Defining System Context using Scenarios. En: Perspectives on Software Requirements, Kluwer Academic Publishers, EEUU, ISBN: 1-4020-7625-8, pp.169-199.

Liu, D., & Shih, T. K. (2003). Transforming natural language requirements into UML diagrams. International Journal of Software Engineering and Knowledge Engineering, 13(4), 379-390.

Maltempo, G., Delle Ville, J., Cecconato, S., Pellegrino, F., Distante, D., Antonelli, L. (2024). Multi-Word Entity Extraction and Rich Relationship Identification to Derive Conceptual Models from Natural Language Specifications. 27th Workshop on Requirements Engineering (WER24), Buenos Aires, Argentina.

Mighetti, J.P., Hadad, G.D.S. (2016). A Requirements Engineering Process Adapted to Global Software Development. CLEI Electronic Journal, Vol. 19, Nº 3, paper 7. ISSN: 0717-5000.

Mira, N.C., Boggio, M.A., Perez, S.B., Salamon, A.G. (2018). Una propuesta de análisis de problemas utilizando LEL y Escenarios en entrenamientos de equipos que gestionan situaciones de crisis, 6to Congreso Nacional de Ingeniería Informática – Sistemas de Información, Mar del Plata, pp. 1041-1048.

Ordinez, L., Donari, D., Santos, R., & Orozco, J. (2010). From User Requirements to Tasks Descriptions in Real-Time Systems. 13th Workshop on Requirements Engineering (WER10), Cuenca, Ecuador, pp. 53-62.

Tjoa, A.M., Berger, L. (1994). Transformation of requirement specifications expressed in natural language into an EER model. En: Elmasri, R.A., Kouramajian, V., Thalheim, B. (eds) Entity-Relationship Approach - ER '93. Lecture Notes in Computer Science, vol. 823. Springer, Berlin.

Zowghi, D., & Gervasi, V. (2002). On the interplay between requirements and architecture: Where do requirements come from and how do they affect software architecture? International Workshop on Software Requirements to Architectures (STRAW).

Técnicas de Control Numérico y Actualización de Software en HPC Aplicadas al Problema de Ríos y Canales Abiertos

Ariel Fraidenraich[1], Francesco Santone[1], Lucas U. Bacchella[1], Facundo H. Bustos[1], Adolfo E. Altenberg[2], Fernando G. Tinetti[3]

[1]Departamento de Ingeniería, Universidad de Belgrano, CABA, Argentina
[2] Departamento de Ingeniería Industrial, UTN-FRBA, CABA, Argentina
[3] III-LIDI, Facultad de Informática, UNLP, La Plata, Argentina
Comisión de Investigaciones Científicas de la Provincia de Buenos Aires

{ariel.fraidenraich, francesco.santone, herrera.facundo, lucas.bacchella}@comunidad.ub.edu.ar, altenberg@gmail.com, fernando@ info.unlp.edu.ar

RESUMEN

Esta línea de investigación tiene dos objetivos fundamentales: 1) Aplicación de técnicas de control de crecidas en flujos de ríos y canales y 2) Aplicación de metodologías estándares de actualización y paralelización de software, incluyendo las conocidas como CI/CD (Continuous Integration/ Continuous Delivery y algunas derivaciones de TDD (Test Driven Development).

CONTEXTO

Para mejorar la resolución de la solución directa se establece un convenio con el Departamento de Estadística de la Universidad Federal de Pernambuco, Recife, Brasil para simular la entrada de una onda de crecida más realista. Se está iniciando un Convenio con la UN-Cuyo, Mendoza para llevar a cabo las mediciones experimentales en obras de riego in situ, y evaluar comprobar el funcionamiento del software desarrollado, especialmente la parte de control y la solución directa (aguas poco profundas). En el esquema de I+D de proyectos de investigación y Desarrollo se presentó un proyecto de investigación en la UB, Universidad de Belgrano, que pudiese describir numéricamente y experimentalmente una entrada de una creciente a un canal o río resueltas por elementos finitos (Taylor - Galerkin) basadas en las ecuaciones de agua poco profundas.

Palabras Claves: FEM (Finite Element Method), Taylor Galerkin, Estadística Hidrológica, Control Experimental, HPC.

1. INTRODUCCION

Se han desarrollado muchas técnicas de control de crecidas basadas en la derivación de caudal hacia un vertedero lateral. las técnicas desarrolladas posteriormente están basadas en la aplicación de metodologías de control tradicional aplicadas a flujo en canales. La entrada de la crecida se simulará con funciones estadísticas [12], [13] en la UFPE (Universidad Federal de Pernambuco), Departamento de Estadística, Recife, Brasil (Convenio UFPE- UB). El enfoque consiste en plantear una función objetivo, definir los parámetros de control y posteriormente hallar la relación entre ambos. Típicamente, la función objetivo vendrá dada en términos del caudal o la altura de la corriente de agua y el parámetro de control es la posición de la compuerta. Una vez que la relación entre

la función objetivo y los parámetros de control está definida, dichos parámetros se pueden modificar para optimizar el objetivo deseado. La desventaja es que el cálculo de la relación entre la función objetivo y los parámetros de control, es decir, las sensibilidades, es un proceso de prueba y error. Esto ha conducido a los investigadores a plantear estrategias de control de resolución más directa. Diversos métodos están basados en la aplicación de esquemas de control óptimo aplicados a formas linealizadas de las ecuaciones que gobiernan el flujo, El camino elegido por estos investigadores, basado en la linealización de las ecuaciones de Saint Venant deja restringida la solución del problema a una clase de flujos limitada. Para evitar esto, Sanders y Katopodes [1], [4], [11] calculan las sensibilidades manteniendo las ecuaciones que gobiernan el flujo en su forma no lineal (ecuaciones de balance de masa y de fuerzas). Este enfoque calcula las sensibilidades necesarias para el control del sistema mediante la resolución de las ecuaciones adjuntas [1], [2], [3]. en aguas poco profundas [7], lo que se denomina adjoint equation method. Esta idea está inspirada en trabajos sobre control óptimo de sistemas debidas a Cacucci [15] y Marchuk [2]. Posteriormente, Hervouet [16], encaró la resolución de problemas similares mediante la utilización códigos abiertos de software. El problema de control se planteará teniendo en cuenta la naturaleza móvil del dominio espacial considerado, es decir utilizando la teoría de contornos móviles [9], [10], [17]. Este trabajo de control optimo se hace optimizando la memoria computacional [5], [6], [8]. Todos los resultados referidos al control óptimo se ensayan experimentalmente en campo en UN-Cuyo (Universidad Nacional de Cuyo) en obras de riego.

2. LINEAS DE INVESTIGACION Y DESARROLLO

Se han desarrollado tareas sobre los temas antes expuestos tales como:

- Definición y evaluación de las mallas de elementos finitos para la simulación numérica
- Función objetivo de la simulación numérica, parámetros de control y evaluación de la relación entre parámetros y la función.
- Tratamiento del sistema de software desarrollado como un sistema de software heredado, de manera tal que se considere como un sistema a mejorar y reestructurar.
- Tratamiento del sistema de software desarrollado como un sistema de cómputo intensivo, para aplicar técnicas de optimización y paralelización estándares en el contexto de los sistemas de HPC (High Performance Computing).
- Aplicación de técnicas de CI/CD y TDD en el desarrollo de las modificaciones del software desarrollado y a actualizar, optimizar y paralelizar.

3. RESULTADOS OBTENIDOS/ESPERADOS

Se implementaron correctamente las ecuaciones de balances del canal. Balance de masa y cantidad de movimiento. Inicialmente, se ha mejorado el programa por medio de paralelización y utilización de técnicas de HPC (computación de alto rendimiento). Se está implementando el problema de control resolviendo la ecuación adjunta y los coeficientes de sensibilidad. Se espera llevar a cabo la verificación experimental en el contexto del convenio entre la UN cuyo -UB (in press). Se espera poder simular de manera más realística por medio de funciones estadís-ticas, como por ejemplo gamas o betas. Se tratará de adicionarle al programa

el problema de bordes móviles. Además, se espera desarrollar un programa que se pueda utilizar en propagación de crecidas. En el 2024 se publica el trabajo "Scientific Legacy Code Enhancement and Testing Strategies" en el Congreso Internacional CSCI 2024 de las Vegas USA (www. american-cse.org/csci2024/) y el trabajo está actualmente en prensa por parte de Springer Nature.

4. Formación de Recursos Humanos

Se formaron tres estudiantes de graduación con su Tesina de grado aprobadas en Ingeniería en informática y en Ingeniería Civil respectivamente. En el año 2025 se anotaron en el Grupo de Investigación tres estudiantes más: uno de Ingeniería Informática de la UB y dos de Ingeniería Civil de la UB todos desarrollando sus respectivas Tesinas de Grado en este marco. Se realizó una colaboración Internacional llamada "Proyecto Columbus", en noviembre del 2024, en la cual uno de los estudiantes participó, generando trabajo relacionado con utilización de alertas y sensores en el proyecto. Entre las tesinas desarrolladas se trabaja en el problema de "dam break" (rotura de presas) y control óptimo. De esta última tesina solo queda la defensa a realizarse en el futuro próximo.

5. Referencias

[1] Sanders, B. F., Katapodes, N. D., "Adjoint Sensitivity Analysis for Shallow Water Wave Control", Journal of Engineering Mechanics, 126 (9), 909-919, 2000.

[2] Marchuk, G. I., Adjoint equations and analysis of complex systems, Kluwer Academic Publishers, Boston, 1995.

[3] Sanders B. F. and Katapodes, D., "Active Flood Hazard Mitigation II Omnidirectional Wave Control ", Journal of Hidraulic Engineering, 125, 1071-1083, 1999.

[4] Sanders B. F. and Katapodes, D., "Control of Canal Flow by Adjoint Sensitivity Method", Journal of Irrigation and Drainage Engineering, 125 (5), 287-297, 1999.

[5] Pacheco P. (2011) An introduction to parallel programming. Morgan Kaufmann, Burlington

[6] Tinetti F. G., Perez M.J., Fraidenraich A., Altenberg A.E. (2018) Experiences in parallelizing a numerical model. In: International Conference on Parallel and Distributed Processing Techniques and Applications, PDPTA'18, Las Vegas, NV, July 30–Aug. 2, ISBN: 1-60132- 487-1, CSREA Press, pp 152–157 26, 2020.

[7] Zienkiewicz OC, Taylor RL (2005) The fnite element method, vol 3, 6th edn. Butterworth-Heinemann, Oxford Publisher's Note

[8] Fernando G. Tinetti; Maximiliano J. Perez; Ariel Fraidenraich; Adolfo E. Altenberg. "Legacy code and parallel computing: updating and parallelizing a numerical model". Journal OF Supercomputing, (2020).

[9] Heniche M., Secretan Y., Baudreau P. and Leclerc M., "A two-dimensional finite element drying-wetting shallow water model for rivers and estuaries", Advances in Water Resources, 23, 359-372, 2000.

[10] Horrit, M. S., "Evaluating wetting algorithms for finite element models of shallow water flow", International Journal for Numerical Methods in Engineering, 55, 835-851, 2001.

[11] Kawahara, M. and Shimada Y.," Gradient Method of Optimal Control Applied to the Operation of a Dam Water Gate", International Journal for Numerical Methods in Fluids.

[12] Ferrari P., Cribari Neto F., "Beta Regression for Modelling Rates", Journal of Applied Statistics, 31, N°7, pp 799-815, 2004.

[13] Bayer F, M., Rosa C. M., Cribari Neto F, "A novel data. Driven dynamic model for inflated double bounded hydro environmental time series, Applied Mathematical Modelling, 137, 2025

[14] Cacuci D. G. Sensitivity and Uncertainty Analysis Theory, volume 1. Chapman & Hall / CRC., 2003.

[15] Hervouet, J. M. Hydrodynamics of Free Surface Flows. Modelling with the Finite element method, 2007.

ORCID autores:
Fernando G. Tinetti:
https://orcid.org/0000-0001-5643-4623

TECNOLOGIAS DE APOYO A LA INFORMÁTICA FORENSE

Liliana Figueroa, Cecilia Lara, Graciela Viaña, Analía Méndez,

Instituto de Investigación en Informática y Sistemas de Información, Facultad de Ciencias Exactas y Tecnologías, Universidad Nacional de Santiago del Estero

sherrera@unse.edu.ar; lmvfigueroa@yahoo.com.ar; laraceciliacristina@gmail.com; gv857@hotmail.com; anmendez725@yahoo.com;

RESUMEN

En la actualidad judicial la implementación del sistema procesal penal acusatorio tiene que ver con un cambio de paradigma, reasignando roles de los actores de la Justicia en el tratamiento de las causas. El sistema acusatorio se llama adversarial, debido al rol que ocupan las partes que intervienen: el juez, la defensa y las fiscalías; en donde recae en los fiscales la tarea de investigar.

En particular, en Santiago del Estero rige la Investigación Penal Preparatoria (IPP), que tiene por objeto determinar la existencia de delitos y la individualización de los eventuales autores; la finalidad es preparar la acusación, determinando las pruebas relevantes que se producirá en el juicio.

En este contexto, el modo de vincular un hecho criminoso con su autor, es a través de pruebas; la mayoría de las IPP cuentan con la asistencia de investigadores y peritos forenses; pero también se destaca que los cambios que se produjeron en estos últimos años donde es necesario encontrar las huellas y pruebas en el ámbito digital.

Entonces el análisis forense informático desempeña un rol decisivo en la resolución de hechos delictivos. Aunque la habilidad y experiencia de los peritos informáticos resulta determinante para este cometido, también resulta importante disponer de herramientas tecnológicas de hardware y software para poder obtener evidencias digitales

En esta propuesta que constituye una continuidad a la línea de investigación de Informática Forense, se pretende investigar sobre herramientas / tecnologías informáticas que permitan gestionar y apoyar de manera eficiente el entorno de los Laboratorios de informática forenses, de manera tal que promueva el trabajo eficiente de la labor de los peritos, ofreciendo un respaldo jurídico a su tarea.

Palabras clave:

Informática forense, evidencias digitales, - herramientas forenses, laboratorios de informática forense.

CONTEXTO

Desde el año 2017 a la fecha el grupo de investigación viene tratando la temática en distintos proyectos de investigación que han sido financiados por el Consejo de Ciencia y Técnica de la Universidad Nacional de Santiago del Estero

En estos proyectos se lograron, hasta la fecha resultados referidos al: análisis de la obtención legal de la evidencia digital en los códigos procesales de nuestro país, análisis de antecedentes jurisprudenciales sobre tratamiento de evidencia digital en dispositivos móviles, investigación y análisis de protocolos vigentes en otras jurisdicciones, propuesta de protocolo para la obtención de evidencia digital móviles en concordancia con las normas la IRAM-ISO/IEC 27037-2022en el ámbito del Ministerio Público Fiscal de Santiago del Estero, y estudio de repositorios que permitan la construcción de un modelo de datos para el almacenamiento y la gestión de evidencias digitales extraídas de dispositivos móviles.

En la implementación del sistema procesal penal acusatorio lo que se busca es ayudar a los fiscales a cumplir con su función y que puedan investigar sin quitarle espacio a la defensa; y que los defensores puedan acceder a toda la causa y que todos los actores de la justicia tengan pleno conocimiento de lo que pasa en el proceso. [1]

En particular, en Santiago del Estero rige el código de procedimiento penal Ley nº 6941 [10], donde la Investigación Penal Preparatoria (IPP), es la etapa del proceso penal que tiene por

objeto determinar la existencia de delitos y la individualización de los eventuales autores; la finalidad es preparar la acusación, determinando las pruebas relevantes que se producirá en el juicio.

En este contexto, el modo de vincular un hecho criminoso con su autor, es a través de pruebas, la ciencia / disciplina encargada de proveer las pruebas son adjetivadas como forenses. En la mayoría de las IPP se cuentan con la asistencia de investigadores y peritos forenses; pero también se destaca que los cambios que se produjeron en estos últimos años donde es necesario encontrar las huellas y pruebas en el ámbito digital. Dispositivos móviles, discos rígidos, correos electrónicos, etc. es donde se debe trabajar actualmente para encontrar las evidencias y pruebas digitales que ayuden a resolver crímenes.

Entonces el análisis forense informático desempeña un rol decisivo en la resolución de hechos delictivos. Aunque la habilidad y experiencia de los peritos informáticos resulta determinante para este cometido, también resulta importante disponer de herramientas tecnológicas de hardware y software para poder obtener evidencias digitales que cumplan los principios de admisibilidad y tengan validez en el proceso judicial [2]. Así un aspecto importante lo constituyen las herramientas de informática forense, las cuales facilitan las tareas de recuperación de archivos y metadatos, la clonación bit por bit de discos duros o el análisis de redes y protocolos.

Considerando que en los últimos años se ha producido un incremento considerable en el volumen de las pericias sobre los dispositivos tecnológicos y que las mismos se presume en mayor escala en el futuro, se plantean algunos de los siguientes inconvenientes:

- El proceso de obtención y análisis de las evidencias digitales es una tarea compleja que demanda un tiempo que en algunos casos excede los plazos procesales.
- No hay un software de informática forense que soporte la extracción de datos de todos los dispositivos móviles existentes en el mundo.
- Costo elevado y disponibilidad presupuestaria limita para la adquisición de software especializado que apoyo las tareas de los peritos.
- Algunos Laboratorios de Informática Forense no cuentan con una estructura funcional y una descripción de los distintos puestos de trabajo, responsabilidades, perfiles para ocupar los puestos de trabajos.
- No hay una cultura en las instituciones de compartir experiencias que se adquieren durante la realización de las pericias y no se fomenta el autoaprendizaje organizacional.

A partir de lo enunciado, en esta propuesta que constituye una continuidad a la línea de investigación de Informática Forense, se pretende investigar sobre herramientas / tecnologías informáticas que permitan gestionar y apoyar de manera eficiente el entorno de los Laboratorios de informática forenses, de manera tal que promueva el trabajo eficiente de la labor de los peritos, ofreciendo un respaldo jurídico a su tarea.

1. INTRODUCCIÓN

La Informática Forense, es una disciplina criminalística, que tiene por objeto, la investigación en sistemas de Tecnologías de Información de hechos con relevancia jurídica o para simple investigación privada. Para conseguir sus objetivos, los expertos se basan en técnicas idóneas para ubicar, reproducir y analizar evidencia digital.

Un principio fundamental en la ciencia forense, empleado continuamente para relacionar un criminal con el crimen que ha cometido, es el "Principio de Intercambio o transferencia", el cual manifiesta que cualquiera o cualquier objeto que entra en la escena del crimen deja un rastro en la escena o en la víctima y viceversa (se lo lleva consigo), en otras palabras: "cada contacto deja un rastro".

Cada crimen tiene una "escena del crimen" que puede llegar a ser asegurada para buscar evidencia; pero algunas veces, la evidencia que se tiene que analizar no es una gota de sangre, una huella digital o la fibra de una alfombra. Son los bits y los bytes contenidos en un dispositivo

que poseen memoria de almacenamiento. [8]

Se entiende por evidencia digital aquellos "registros que fueron procesados en un dispositivo informático y se encuentran almacenados o fueron transmitidos a través de un medio de comunicación informático" o "toda información electrónica que puede ser utilizada como fuente de prueba en un procedimiento judicial ante allanamientos tradicionales o remotos".

La evidencia digital posee, por su naturaleza, características específicas, en primer lugar "latentes", esto es, que la información que las compone se encuentra contenida en un soporte físico (pendrive, memoria RAM de un dispositivo, etc.) que el ser humano no puede visualizar sin la ayuda de procedimientos matemáticos complejos. En segundo lugar, "frágiles y volátiles", en el sentido que pueden ser alteradas o dañadas fácilmente y sin que ello deje rastro alguno. Por último, poseen la "capacidad de transferirse transfronterizamente", ya que, a través de Internet, una evidencia digital puede disiparse en diferentes partes del mundo. [11]

Según [12], la evidencia digital contiene algunas características propias que la distinguen de cualquier otra evidencia tradicional. Entre otras, contiene las siguientes particularidades:

- Volátil: Puede perderse si no se recolecta en tiempo y forma.
- Duplicable: Pueden realizarse diversas copias sin poder reconocer el original.
- Alterable: Factible de su modificación y/o borrado y sin el registro de esas acciones.
- Anónima: En algunos casos no se pueden determinar el autor de las mismas.

Posee datos adicionales no visibles por las herramientas tradicionales usadas por el usuario: Último acceso, modelos de cámara usados en una fotografía, coordenadas de geo posicionamiento, autor, última impresión, secuencia de recorrido de un correo electrónico, etc.).

La Informática forense usa la evidencia digital como elemento material probatorio en un proceso judicial [4] [3]; de esta manera constituye una disciplina auxiliar a la justicia, que consiste en la aplicación de técnicas que permiten adquirir, validar, analizar y presentar datos que han sido procesados electrónicamente y guardados en un medio computacional. [5]

Por otra parte, el FBI, considera que la Informática forense es la ciencia que se encarga de adquirir, preservar, analizar y presentar los datos que han sido procesados electrónicamente y almacenados en medios electrónicos aplicando técnicas científicas y analíticas, utilizando hardware y software especializado para realizar la tarea. [9]

El análisis forense es una disciplina que permite identificar, analizar, preservar y presentar evidencia digital obtenida de infraestructura tecnológica de tal manera que sea válida en un proceso legal. Esta infraestructura tecnológica puede ser desde un teléfono celular, una cámara digital, una computadora, una impresora o una memoria digital, es decir, elementos que tienen memoria. [8]

En este contexto los elementos informáticos que se disponen en los celulares, las computadoras o los dispositivos de almacenamiento como imágenes, mensajes de texto, conversaciones de WhatsApp, videos, ubicación, etc. constituyen las evidencias necesarias para resolver diferentes tipos de casos judiciales. Todos estos elementos son trabajados en los Laboratorios de informática forense, en donde se realizan las extracciones forenses apoyadas con hardware y software especializado; luego la información obtenida es procesada y analizada para conseguir la evidencia digital necesaria para las investigaciones. Por otra parte, en estos laboratorios no solo hay que recuperar la información, sino también realizar procedimientos de la forma adecuada para poder utilizar esta información como evidencia; entonces hay que garantizar la confiabilidad de la prueba y la cadena de custodia de la información, cuando se realiza una operatoria de estas características.

En [7], se presenta una propuesta de los procedimientos para el establecimiento y gestión de Laboratorios informática forense, proporcionando técnicas para el manejo y

procesamiento de pruebas digitales. Tienen como objetivo proporcionar un marco universal para establecer y administrar un Laboratorio de informática forense. Para recoger y manipular pruebas digitales, se debe tener en cuenta el cumplimiento de principios básicos definidos por [6]. Estos principios sirven como marco de referencia y apoyan los procedimientos de actuación que se desarrollen en dichos laboratorios.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Considerando la amplitud de los aspectos relacionados con la informática forense, en el presente proyecto de investigación se diferencian las siguientes líneas de investigación:

- Estudio y Desarrollo de herramientas tecnológicas para el apoyo de la Gestión de los Laboratorios de informática forense.
- Técnicas y métodos para la gestión de laboratorios de informática forense.

3. OBJETIVOS

El objetivo general de la investigación propuesta es: *Contribuir a la eficiencia y calidad de la gestión de evidencias digitales en el contexto del sistema judicial santiagueño.*

Los objetivos específicos por cada línea de investigación que permitirán alcanzar el objetivo general son:

LI1: Estudio y Desarrollo de herramientas tecnológicas para el apoyo de la Gestión de los Laboratorios de informática forense.

a. Estudiar sistemáticamente herramientas de informática forense.

b. Desarrollar aplicaciones que permitan dar apoyo a las tareas del perito informático.

c. Desarrollar un prototipo de repositorio privado de evidencias digitales, permitiendo la gestión centralizada, eficiente y segura de la evidencia digital capturada

LI2: Técnicas y métodos para la gestión de laboratorios de informática forense.

d. Evaluar y mejorar los procesos internos de los laboratorios de informática forense para garantizar el cumplimiento de estándares de calidad y normativas legales

Se trata de una investigación aplicada, desarrollada desde un enfoque cuanti-cualitativo, en el campo de la Informática Forense.

La investigación es aplicada porque los resultados que se obtengan se aplicarán inmediatamente la oficina de Informática Forense del Gabinete de Ciencias Forenses del Ministerio Público Fiscal, en el Gabinete de Informática Forense del Poder Judicial de la Provincia de Santiago del Estero, en la Policía de la provincia de Santiago del Estero, en la Oficina de Informática del Ministerio Publico de la Defensa.

4. FORMACIÓN DE RECURSOS HUMANOS

El grupo de investigación está conformado principalmente por el grupo que ya viene trabajando en la línea de informática forense desde el año 2017 en los proyectos ya mencionado, quienes ya tienen antecedentes: MSc. Ing. Liliana Figueroa, Dra. Graciela Viaña de Avendaño, Licenciada Cecilia Lara, Ing. Analía Méndez.

Para trabajar en el desarrollo de aplicaciones que mejoren la funcionalidad de las herramientas forenses y el desarrollo de soluciones informáticas para la gestión eficiente de los laboratorios forenses se incorporan a la Profesora Claudia Gómez, Licenciadas: María de los Ángeles Valdez y Fernanda Leguizamón

La participación de profesionales vinculados a las ciencias forenses, de egresados y estudiantes avanzados vinculados con los distintos organismos judiciales del nuevo sistema penal procesal posibilitará concretar las pruebas e implementaciones de los resultados de la investigación.

5. REFERENCIAS

[1] Argentina.gob.ar. (2021, 6 de septiembre). Sistema acusatorio: nuevos roles y mayor eficacia. https://www.argentina.gob.ar/noticias/sistema

-acusatorio-nuevos-roles-y-mayor-eficacia

[2] Calderón Valdiviezo RG, Guzmán Reyes GS, Salinas González JM, Aranda A. Diseño y Plan de Implementación de un Laboratorio de Ciencias Forenses Digitales. (2011). file:///C:/Users/hp/Downloads/paper_laboratorio_forense_digital.pdf

[3] Castillo, C., Romero, A., Cano, J.: Análisis Forense Orientado a Incidentes en Teléfonos Celulares GSM: Una Guía Metodológica. Conf. XXXIV Conferencia Latinoamericana de Informática, Centro Latinoamericano de Estudios en Informática (CLEI). (2008). http://www.clei2008.org.ar/

[4] Del Pino, S.: Introducción a la informática forense. Pontificia Universidad Católica del Ecuador. (2007). http://www.alfa-redi.org/sites/default/files/articles/files/Acurio.pdf

[5] Di Iorio, Ana H. -La informática forense y el Proceso de recuperación de Información digital. Revista Democracia Digital e Governo Eletrônico (ISSN 2175-9391), n° 8, p. 326-339 - 2013.

[6] Digital Evidence: standards and principles", Forensic Science Communications, FBI, apr. 2000. https://archives.fbi.gov/archives/about-us/lab/forensic-science-communications/fsc/april2000/swgde.htm

[7] Directrices Mundial de INTERPOL para los Laboratorios Forenses Digitales. INTERPOL L Complejo Mundial para la Innovación de 2019. htps/Downloads/INTERPOL_DFL_GlobalGuidelinesDigitalForensicsLaboratory%20(2).pdf

[8] Iglesias LR. Herramientas open source para Informática forense. Carrera de Especialización en Seguridad Informática, Facultades de Ciencias Económicas, Ciencias, Exactas y Naturales e Ingeniería, Universidad de Buenos Aires. (2015). http://bibliotecadigital.econ.uba.ar/download/tpos/1502-0187_IglesiasLF.pdf

[9] Noblett M. G., Pollit, M. M. "FBI" FBI, October 2000. https://archives.fbi.gov/archives/about-us/lab/forensic-science-communications/fsc/oct2000/computer.htm

[10] Ley provincial 6.941. Código Procesal Penal de la Provincia de Santiago del Estero (2009).

[11] Sassón E, Arroyo M. Donnangelo AP. Cadena de custodia. El debido tratamiento de la evidencia digital en el proceso penal. Revista Uruguaya De Derecho Procesal (2021). Disponible en: https://revistas.fcu.edu.uy/index.php/rudp/article/view/4619/4043

[12] Sosa ME. Revista Pensamiento Penal (ISSN 1853-4554), No. 466. Mayo de 2023. Disponible en: https://www.pensamientopenal.com.ar/doctrina/90772-evidencia-digital-su-importancia-investigacion

UN MODELO DE CALIDAD CON BASE EN LA AGILIDAD PARA ALCANZAR LA TRANSFORMACIÓN DIGITAL

Mario Peralta[1], Carlos Salgado[1], Daniel Riesco[1], Lorena Baigorria[1], Javier Saldarini[2], Claudio Carrizo[2], Germán Montejano[1]

[1] Departamento de Informática, Facultad de Ciencias Físico-Matemáticas y Naturales
Universidad Nacional de San Luis, Ejército de los Andes 950, San Luis
{mperalta, csalgado, driesco, flbaigor, gmonte}@unsl.edu.ar
[2] Facultad Regional San Francisco, Universidad Tecnológica Nacional
{saldarinijavier, cjcarrizo77}@gmail.com

RESUMEN

A medida que el comercio digital se instala entre los consumidores, crece la necesidad de captar a esos usuarios. La demanda de técnicas, herramientas y estrategias crece al ritmo que se multiplican los canales digitales de venta y de cómo los consumidores siguen adoptando nuevos hábitos de compras online. Es por esto que los especialistas en ecommerce hoy se posicionan como los perfiles más demandados del área comercial.

Debido al alto índice de crecimiento del ecommerce a nivel global, las empresas que no contaban con conocimientos acerca de este canal de comercialización, tuvieron que adecuarse para continuar con su negocio de manera competitiva.

Para ello se propone usar la agilidad como un hilo conductor para defini**r un modelo de calidad workflow**, métricas e indicadores basado en estándares de calidad e información recabada de la industria del software y servicios informáticos de la región, haciendo hincapié en una arquitectura de integración con otros sistemas para concretizar la transformación digital de las empresas.

Palabras Clave: Agilidad. Modelo de Calidad. Workflow. Transformación Digital.

CONTEXTO

El presente trabajo se enmarca en el Proyecto de Investigación: Ingeniería de Software: Estrategias de Desarrollo, Mantenimiento y Migración de Sistemas en la Nube – Facultad de Ciencias Físico-Matemáticas y Naturales, Universidad Nacional de San Luis. Proyecto Nº P-03-2020. Dicho proyecto es la continuación de diferentes proyectos de investigación a través de los cuales se ha logrado un importante vínculo con distintas universidades a nivel nacional e internacional. Además, se encuentra reconocido por el programa de Incentivos.

1. INTRODUCCIÓN

La búsqueda de calidad en los productos de software data de mucho tiempo atrás, desde los modelos propuesto por McCall (1976) [1], donde se descompone la calidad del producto en 11 categorías o factores de calidad agrupándolos en tres categorías: operación, revisión y transición de productos; pasando por Evans (1987) [2], Deustch (1988) [3], que agregaron nuevas categorías y factores de calidad a los ya propuestos. Estos modelos fueron desarrollados siguiendo lineamientos de calidad que surgieron para la mejora de otros tipos de productos [4, 5]. En la actualidad, también fueron evolucionando distintas familias de normas de calidad orientadas a productos de software. Las normas ISO/IEC 9126 [6] y ISO/IEC 14598 [7] dieron forma a la familia de ISO/IEC 25000, familia de normas también conocida como SQuaRE (Software Product Quality Requirente and Evaluation) [8]. Es así como la familia de normas 25000 se organiza en cinco puntos: requisitos, gestión, modelo, medición y evaluación de la calidad, se exige en forma contractual aquella correspondiente al modelo de calidad -ISO/IEC

25010 [9]. La norma propone a su vez tres modelos: modelo de calidad del producto de software, modelo de calidad en el uso del sistema y modelo de calidad de datos. La norma ISO 25010 divide a la calidad del producto de software en 8 características: Funcionalidad, Fiabilidad, Eficiencia, Usabilidad, Seguridad, Compatibilidad, Mantenibilidad y Portabilidad.

Además, se debe tener presente también los modelos de negocio. En la actualidad, los modelos de negocio **B2B** (Business to Business), **B2C** (Business to Consumer), **B2B2C** (Business to Business to Consumer) están migrando a un modelo **D2C** (Direct to Consumer) en el cual se hace hincapié en el contacto con los consumidores. Entonces, tienen que integrar toda la cadena de valor y la propuesta para hablarle a quien adquiere sus productos, aunque ellos no los vendan directamente. El escenario se ha modificado, así, desarrollar un producto digital sin preguntarnos quiénes lo usarán, para qué, desde qué dispositivos, etc., no hace más que alejarnos de la premisa que reina hoy en día en esta era digital: *todo por el usuario*.

A menudo en la industria del software nos encontramos con productos digitales que son ideados y desarrollados por personas expertas en el tema, a las que les resulta fácil entender el uso de un *menú hamburguesa* o que conviven en lo cotidiano con términos como *scroll, input* o *form.* Pero ¿qué pasa si hablamos de usuarios que no son expertos en el tema? ¿Y si hablamos de los diferentes grupos etarios? ¿De comunidades vulnerables o que no poseen fácil acceso a la tecnología? ¿Cómo se sienten al encontrarse con un producto digital que no fue ideado o desarrollado para su situación o su contexto?

Ahora preguntémonos: ¿cómo podrían los equipos de desarrollo sacar al mercado un producto digital pensando en el usuario si generalmente trabajan contra reloj? Si a menudo conocen al usuario en el transcurso del desarrollo, ¿podrían cubrir las modificaciones que van siendo solicitadas? A las respuestas sumémosle que los equipos están conformados por personas con diferente visión, y que pueden pensar diversas soluciones según su experiencia ¿Cómo se podría lograr un consenso?

Y si tenemos en cuenta que, a menudo, los equipos de desarrollo deben producir software multiplataforma, impactando directamente sobre la experiencia del usuario (UX: User eXperience) por los riesgos propios de las tecnologías que se usan [10], ¿podríamos llegar a un producto digital con el que todos, equipo y usuarios, se sintieran satisfechos e identificados?

Luego de conocer las consideraciones obtenidas en la investigación y de analizar la perspectiva de diferentes autores, podemos comprender que, conocer el contexto y a los usuarios que harán uso de nuestro producto/servicio nos dará la posibilidad de delimitar nuestro campo de diseño y desarrollo, optimizando esfuerzos y recursos, para obtener un producto de calidad asegurando mejorar la relación usuario-producto.

Así, la innovación puede verse como una forma de perdurar en el tiempo. El éxito del proceso de innovación tecnológica en toda organización/empresa involucra a todos los integrantes de la misma. Desde los ejecutivos que toman las decisiones que impactan en toda la cadena de valor, hasta los colaboradores que trabajan todos los días para lograr el producto final que llega a los consumidores.

Estos cambios no solo hacen más ágil la operación de la compañía, sino que les permite responder ante las tendencias y necesidades del mercado de manera mucho más rápida.

La calidad de un producto de software desde el punto de vista del cliente, se define como “el grado en que un cliente y/o usuario percibe que el producto de software satisface sus necesidades” [8]. Bajo este concepto se establecen las normas de calidad. En el año 2005 surge el estándar IOS/IEC 25000 [11] que permite medir la calidad del software en sí mismo; y si bien la norma permite controlar y evaluar la calidad de un producto, los costos de su certificación e implementación resultan elevados para pequeñas empresas de software.

El presente trabajo introduce un modelo de calidad que cubre lo referente a UX y satisface la necesidad de adaptarse a los nuevos paradigmas de negocio, considerando los

cambios tecnológicos y de mercado actuales. En definitiva, ofrecer un instrumento que le permita conocer el dominio con el que trabaja, familiarizarse con el vocabulario propio de la organización y establecer las herramientas, plataformas, técnicas y métodos de comunicación necesarios para lograr un proceso que satisfaga las necesidades, tanto de la empresa como del cliente.

2. PROPUESTA

Cada vez más en el último tiempo, y especialmente a causa de la pandemia y los avances tecnológicos, la tendencia de la transformación digital encarada por las empresas fue creciendo y se instaló como tal, siendo un proceso hoy en marcha en mayor o menor medida por gran parte de las compañías. Esto les Permite posicionarse en el mercado y ser más competitivas.

Usualmente lo primero que se piensa está relacionado con la tecnología y, sin embargo, la transformación digital es mucho más. Hoy se vive en un mundo en continuo cambio y disrupción. Así, Hay tres aspectos que permiten visualizar esos cambios:

I. Lo que se consume. (Lo que nos interesa).
II. Cómo se lo consume. (Cómo lo conseguimos o nos apropiamos).
III. Cómo se lo comunica. (cómo nos enteramos o conocemos).

Esto significa para las organizaciones que se abren muchísimas oportunidades que se hacen viables gracias a la tecnología.

3. ÁRBOL DE REQUERIMIENTOS

El siguiente árbol de requerimientos resume un conjunto de características y subcaracterísticas que son deseables, como mínimo, para hablar de transformación digital:

1. Mejorar las decisiones
   1.1. Datos
   1.2. Información
   1.3. Conocimiento
2. Mejorar la Experiencia
   2.1. Del Producto
   2.2. Del Servicio
   2.3. De los Clientes
      2.3.1. Fidelización
      2.3.2. Retención
3. Reducir Riesgos
   3.1. Fallos de Automatización
   3.2. Fallos de Recursos / Dispositivos
4. Trabajo Colaborativo
   4.1. Uso Plataformas Digitales
   4.2. Trabajo en Equipo
      4.2.1. Diferentes Áreas
      4.2.2. Diferentes Lugares Geográficos
5. Incrementar la Productividad
   5.1. Herramientas
   5.2. Aplicaciones
6. Reaccionar al cambio
7. Generar Aprendizajes y Adaptación
   7.1. Personas
      7.1.1. Curiosidad
      7.1.2. Creatividad
      7.1.3. Empatía
      7.1.4. Colaboración
   7.2. Organización
      7.2.1. Curiosidad
      7.2.2. Creatividad
      7.2.3. Empatía
      7.2.4. Colaboración
8. Foco en el Cliente
   8.1. Necesidades
   8.2. Visión del Cliente
9. Promover la Innovación
10. Trabajar diferenciado
   10.1. Trabajo Colaborativo
   10.2. Auto Gestión

A continuación, se da una breve definición de las características y subcaracterísticas del modelo de calidad (representadas en al árbol de requerimientos presentado) para llevar a cabo una transformación digital

1. Mejorar la toma de decisiones basada en datos, información y/o conocimiento, siendo éste el punto de partida para el planeamiento estratégico identificando nuevas oportunidades de negocio.
2. Mejorar la experiencia de nuestros clientes. Ya no alcanza con el producto o el servicio, sino que desea recibir una

atención diferencial, estrechamente relacionada con la fidelización y retención de clientes.
3. Reducir riesgos y fallas con la automatización de procesos de negocio.
4. Fortalecer el trabajo colaborativo utilizando plataformas digitales, lo que da lugar a que la comunicación entre cada uno de ellos sea más fluida, así como el trabajo en equipo entre las diferentes áreas.
5. Aumentar la productividad utilizando herramientas y aplicaciones digitales corporativas.
6. Reaccionar al cambio cada vez más rápido y así diferenciarnos de nuestros competidores.
7. Generar la capacidad de adaptación y aprendizaje: curiosidad, creatividad, empatía, colaboración, en las personas y en la organización
8. Enfocarse en el cliente, comprender sus necesidades, ver el mundo desde el punto de vista de nuestros clientes.
9. Promover la innovación, como pilar fundamental en al cambio cultural de las organizaciones.
10. Hacer las cosas diferentes: trabajo colaborativo, empoderamiento y auto gestión.

Para lograr llevar acabo la transformación digital, es necesario e indispensable abordar la agilidad. Se puede decir que la agilidad es un catalizador de la transformación digital. Podríamos pensarla como la habilidad de adaptación al cambio, un *mindset* que permite abordar la complejidad de una manera adaptativa para responder rápidamente a las necesidades del negocio. *Mindset* es un término ampliamente utilizado en el mundo del emprendimiento. El mindset es el conjunto de creencias, actitudes y patrones de pensamiento que influyen en nuestra forma de percibir y responder a situaciones y desafíos [12].

Así, es importante tener claro qué es y qué no es la agilidad:

- No es velocidad: es adaptación al aprendizaje, es foco.
- No es baja calidad: es calidad como pre requisito.
- No es una meta: es un camino.
- No es una solución: es una alternativa a problemas complejos.
- No es una metodología: es un *mindset*.

Podemos decir que la agilidad promete colaboración, entrega iterativa e incremental, adaptación al cambio y comunicación.

En resumen, si se buscan resultados diferentes, se necesita hacer algo diferente y en ese caso el modelo de calidad definido nos marca el nuevo camino para ganar agilidad en el proceso de la transformación digital.

4. RESULTADOS OBTENIDOS Y OBJETIVOS

En base a lo expresado, y el Árbol de requerimientos presentado, se definió un modelo de calidad que permitiera pensar la transformación digital con base en la agilidad. Proponiendo un conjunto de métricas e indicadores asociadas a las características y subcaracterísticas listadas en el Árbol de requerimientos.

En el ámbito de nuestro trabajo de investigación, hemos definido un conjunto de métricas para la medición de la calidad de modelos de procesos workflow desde el punto de vista de su mantenibilidad [13].

Los principales ejes de trabajo de esta línea de investigación se detallan a continuación:

- Realizar una Revisión Sistemática de la Literatura (RSL) enfocada en los aspectos relacionados con la calidad de productos de software/datos y su aplicación, para llevar a cabo especificaciones de requisitos de calidad.
- Realizar un estudio en profundidad de la Familia de Normas ISO/IEC 25000, en particular de las Normas ISO/IEC 25030, ISO/IEC 25010, ISO/IEC 25012, ISO/IEC 25023 e ISO/IEC 25024.
- Realizar un estudio del marco de trabajo ágil propuesto por Scrum y su aplicación en el desarrollo de software.

- Identificar de qué manera se realiza la identificación, selección y especificación de requisitos de calidad de la empresa, clientes, software y de datos, en empresas que estén en camino a una transformación digital.
- Definir como propuesta un modelo/método/técnica que facilite la incorporación de requisitos de calidad para la transformación digital.
- Validar dicha propuesta en casos de estudios reales, a fin de poder realizar una contratación, que permita realizar los ajustes correspondientes.

5. FORMACIÓN DE RECURSOS HUMANOS

Dentro del proyecto, donde se encuentra inserta esta línea de investigación, se están llevando a cabo diferentes tesis de grado y posgrado. Éstas últimas correspondientes a la Especialización en Ingeniería de Software, Maestría en Ingeniería de Software, Maestría en Calidad del Software y al Doctorado en Ingeniería Informática. En particular, la línea de investigación forma parte de una tesis del Doctorado en Ingeniería informática.

BIBLIOGRAFÍA

[1] J. A. McCall and P. K. Richards, "Factors in software quality. US Rome Air Development Center Reports," 1977.

[2] M. W. Evans and J. J. Marciniak, "Software Quality Assurance and Management. New York: Jhon Willey &Sons.," 1987.

[3] M. S. Deustch and R. R. Willis, "Software Quality Engineering, A Total Technical Management Approach. Prentice Hall.," 1988.

[4] M. Sangüesa, R. Mateo, and L. Hzarbe, "Teoría y Práctica de la Calidad. ," *España: Thomson.,* 2006.

[5] M. G. Piattini Velthuis, F. O. García Rubio, I. García Rodriguez de Guzmán, and F. Pino, *Calidad de Sistemas de Información*, 2012.

[6] ISO/IEC, "ISO/IEC 9126-1:2001. Software Engineering, "Software Product Quality Part 1: Quality Model, Int'l Org. For Standardization", Geneva, Switzerland," ed, 2001.

[7] ISO/IEC, "ISO/IEC 14598-1. Information Technology "Software Product Evaluation Part 1: General overview"," ed, 1999.

[8] ISO, "ISO 9241-210:2010 Ergonomics of human-system interaction -- Part 210: Human-centred design for interactive systems," ed, 2010.

[9] ISO, "ISO/IEC 25010:2011 Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- System and software quality models," ed, 2011.

[10] A. Biørn-Hansen, T.-M. Grønli, G. Ghinea, and S. Alouneh, "An Empirical Study of Cross-Platform Mobile," *Wireless Communications and Mobile Computing, vol. 2019, nº 5743892, 3,* 2019.

[11] ISO, "ISO/IEC 25000. SQuaRE - System and Software Quality Requirements and Evaluation.," ed.

[12] C. Dweck, *MINDSET LA ACTITUD DEL ÉXITO*, Spanish Edition ed.: Editorial Sirio, 2016.

[13] M. Peralta, D. Riesco, C. Salgado, G. Montejano, and F. Ó. Garcia Rubio, "Un Conjunto de Medidas para la Evaluación de Modelos Workflow," in *Congreso Argentino en Ciencias de la Computación - CACIC 2008*, 2008.

VERIFICACIÓN Y VALIDACIÓN DE SOFTWARE EN LA INDUSTRIA 5.0

Martín L. Larrea[1,3,4], Dana K. Urribarri[1,3,4], M. Luján Ganuza[1,3,4], Matías N. Selzer[1,3,4], M. Laura Cobo[1], Antonella S. Antonini[1,3,4], Leila Vecslir[2]

{mll, dku, mlg, matias.selzer, mlc}@cs.uns.edu.ar, antoo.antonini@gmail.com, leila.vecslir@uns.edu.ar

[1]Departamento de Ciencias e Ingeniería de la Computación

[2]Departamento de Economía

[3]Laboratorio de I+D en Visualización y Computación Gráfica (VyGLab) (UNS-CIC)

[4]ICIC, Instituto de Ciencias e Ingeniería de la Computación (UNS-CONICET)

Universidad Nacional del Sur

RESUMEN

La Industria 5.0 amplía el enfoque de la Industria 4.0 al reconocer el potencial de la industria para alcanzar objetivos sociales que van más allá del empleo y el crecimiento económico, colocando el bienestar del trabajador en el centro del proceso productivo. A partir del estado del arte en verificación y validación de software, así como de los principios de la Industria 5.0 en la interacción entre usuarios y software, es posible diseñar, implementar y adaptar metodologías y herramientas que faciliten la detección de errores y problemas en los factores de calidad, tanto internos como externos, del software. Este trabajo de investigación tiene como propósito el diseño y desarrollo de herramientas que contribuyan a garantizar la calidad del software, con especial énfasis en la satisfacción del usuario durante su uso. Además, el proyecto contempla la creación de espacios de formación dirigidos a trabajadores sin formación informática, brindándoles capacitación para llevar a cabo la validación del software en sus aspectos funcionales y no funcionales. De esta manera, se promueve la mejora en la capacitación de los recursos humanos y un mayor control sobre la calidad del software.

1. INTRODUCCIÓN

En el año 2021 la Comisión Europea identificó formalmente a la Quinta Revolución Industrial (denominada Industria 5.0), tras debates entre participantes de organizaciones de investigación y tecnología, y otras agencias de toda Europa. Esto quedó reflejado en el reporte producido por la comisión "Industry 5.0 - Towards a sustainable, human-centric and resilient European industry" [1].

La Industria 5.0 complementa el paradigma existente de la "Industria 4.0" al reconocer el poder de la industria para lograr objetivos sociales más allá del empleo y el crecimiento, para convertirse en un proveedor resiliente de prosperidad, haciendo que la producción respete los límites de nuestro planeta y colocando el bienestar del trabajador de la industria en el centro del proceso de producción. La investigación y la innovación se centran en impulsar la transición hacia una industria sostenible, centrada en las personas. La Quinta Revolución Industrial mantiene al software como elemento transversal y fundacional, en especial a través de la reconciliación que pretende entre humanos y máquinas [2]. Las metas que se proponen, específicamente aquellas vinculadas a la relación entre los usuarios y el software, redefinen lo que es nuestro entendimiento de la calidad de los sistemas informáticos. En tal sentido, y en sintonía con los objetivos planteados, la experiencia del usuario y su satisfacción se han transformado en las principales metas de calidad [3]. Los equipos interdisciplinarios son una forma cada vez más común de desarrollar software, por lo que las herramientas que garantizan la calidad de los productos de software deben ser utilizables por todos los miembros del equipo, y no sólo por aquellos con un perfil informático [4].

Esta nueva era requiere nuevas metodologías, técnicas y herramientas para verificar y validar el software. La ingeniería busca la

producción de software de calidad, entendida como una combinación de varios factores [5]. Por un lado, consideramos cualidades como la facilidad de uso, la adaptabilidad y la accesibilidad cuya presencia o ausencia en un producto de software puede ser detectada por sus usuarios. Estas propiedades se denominan factores de calidad externos. Otras cualidades aplicables a un producto de software, como ser modular o extensible, son factores internos, perceptibles sólo para los profesionales de la informática que tienen acceso al código del software.

La Industria 5.0 pone la atención en los factores externos. Sin embargo la clave para lograr estos factores externos está fuertemente vinculada con los internos: para que los usuarios disfruten de las cualidades visibles, los diseñadores y desarrolladores deben haber aplicado técnicas internas que aseguren las cualidades ocultas. Para poder potenciar ambas caras de la calidad, debemos introducir nuevas formas de guiar el desarrollo de software y de también evaluar lo ya desarrollado.

La Verificación y la Validación son dos formas de evaluar la calidad de un producto de software. La Verificación se ocupa de comprobar que el software cumple los requisitos funcionales y no funcionales de su especificación, los factores internos; mientras que la Validación tiene a su cargo comprobar que el software cumple las expectativas que el cliente espera, los factores externos.

En este contexto industrial, para avanzar en software con un enfoque centrado en las personas es necesario repensar y rediseñar las formas de desarrollar y evaluar los sistemas. Los usuarios deben participar en cada paso de estos procesos dando retroalimentación. Nuestra realidad actual nos muestra que cada día más decisiones son tomadas en base a información brindada por un sistema informático, como es el caso del análisis visual mediante el uso de representaciones visuales de grandes conjuntos de datos, Realidad Virtual o Aumentada. No sólo hay un aumento cuantitativo sino también cualitativo de estas decisiones, las mismas son más críticas y con mayor impacto en la sociedad, el medio ambiente y los individuos.

En el nivel de implementación, debemos asegurar la calidad de los factores internos mediante la verificación del software. Los nuevos desafíos que se enfrentan conllevan nuevas aplicaciones de los sistemas informáticos y nuevos paradigmas de desarrollo, en consecuencia, se requieren adaptar las formas de encontrar errores o generar nuevas técnicas de testeo. Desde la perspectiva del usuario, debemos asegurar la calidad de los factores externos mediante la validación del software. La meta que establece la Industria 5.0 de ubicar el factor humano en el centro de la producción nos obliga a trasladar esto al software involucrado en dicha producción. El factor humano se presenta en la Ingeniería de Software a través de las áreas de Usabilidad e Interacción Humano-Computadora; áreas que existen desde hace décadas y tienen mucho para aportar en esta nueva revolución industrial.

Esta línea tiene como finalidad desarrollar nuevas técnicas de verificación y validación de software en sintonía con la Industria 5.0. Además, se busca fomentar e instruir a aquellos perfiles de trabajo no vinculados con los sistemas informáticos en aspectos relacionados con el testeo de software y el diseño centrado en el usuario.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Basándonos en el estado del arte de la verificación y validación de software, y en los objetivos establecidos para la Industria 5.0 en cuanto a la relación usuario/software, es posible definir, implementar y adaptar metodologías y herramientas que permitan detectar errores y/o problemas, tanto en los factores de calidad externos y/o internos del software.

La generación de espacios de formación orientados a perfiles de trabajadores no

informáticos, permite su capacitación para poder realizar la validación de software en aspectos funcionales y no funcionales. Esto logra mejorar la capacitación de los recursos humanos y un mayor control de la calidad del software.

El objetivo general consiste en el estudio, desarrollo y difusión de metodologías para la validación y verificación de software en el contexto de Industria 5.0. El objetivo específico de este trabajo de investigación consiste en el diseño y desarrollo de herramientas que asistan en el desarrollo software para asegurar la calidad del producto final, particularmente en aquellos aspectos vinculados con la satisfacción del usuario al usar el software. También es un objetivo específico el de generar capacitaciones a perfiles no técnicos sobre cómo usar estas herramientas.

3. RESULTADOS OBTENIDOS/ESPERADOS

Esta línea de investigación es resultado de diversos trabajos publicados cuyos resultados se resumen en [6].

El equipo se encuentra trabajando en expandir las técnicas presentadas en [6] para incluir una estrategia de testing orientado a objetos a nivel de clase con una metodología modal [8].

A partir de esta línea de trabajo se espera expandir las metodologías de testing a nivel de experiencias de usuario y de funcionalidad.

4. FORMACIÓN DE RECURSOS HUMANOS

En lo concerniente a la formación de recursos humanos se detallan las tesis en desarrollo y concluidas relaciones con las líneas de investigación presentadas, así también como un proyecto de grupo de investigación.

Tesis de Grado

- *Trayecto Formativo en Verificación y Validación de Software Dentro de una Empresa de Desarrollo de Software.* Ezequiel Mauricio Zucoli. 2024
- *Verificación y Validación de Software SwipTesting*, Nicolás Martín Leidi, 2024.

Proyecto de Grupo de Investigación

- Verificación y Validación de Software en la Industria 5.0, PGI 2024-2028.

Al momento de la presentación de este trabajo se cuenta con dos tesis de grado en desarrollo.

5. BIBLIOGRAFÍA

[1] European Commission, Directorate-General for Research and Innovation, Breque, M., De Nul, L., Petridis, A., Industry 5.0 – Towards a sustainable, human-centric and resilient European industry, Publications Office of the European Union, 2021, https://data.europa.eu/doi/10.2777/308407

[2] Akundi, Aditya, et al. "State of Industry 5.0—Analysis and identification of current research trends." Applied System Innovation 5.1 (2022): 27.

[3] Riasat, H. A. F. S. A., et al. "Enhancing Software Quality Through Usability Experience And HCI Design Principles." vol 42 (2023): 46-75.

[4] Ustundag, Alp, and Emre Cevikcan. Industry 4.0: managing the digital transformation. Springer, 2017.

[5] Jena, Divya, et al. "Systematic Literature Review on Object Oriented Software Testing Techniques." 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA). IEEE, 2023.

[6] Larrea, Martín L., and Dana K. Urribarri. "New verification and validation tools for Industry 4.0 software." Designing Smart Manufacturing Systems. Academic Press, 2023. 61-88.

[7] Larrea, Martín Leonardo, et al. "Verificación y validación de

Visualizaciones." XXV Workshop de Investigadores en Ciencias de la Computación (WICC 2023, Junin, Buenos Aires). 2023.

[8] Binder, Robert V. "Testing object‐oriented software: a survey." Software Testing, Verification and Reliability 6.3‐4. 1996. 125-252.

· IEI ·

Innovación en Educación en Informática

Cluster Cronos Mejorado: Expansión y Optimización

Martha Semken, Mariano Vargas, Ignacio Tula, Giuliana Zorzoli ,Andrés Rojas Paredes

Instituto de Ciencias (ICI) Universidad Nacional de General Sarmiento – UNGS

{msemken, avargas, itula, gjzorzoli ,arojas}@campus.ungs.edu.ar

RESUMEN

El cómputo de alto rendimiento (HPC, por sus siglas en inglés) es una herramienta clave para la resolución de problemas complejos en diversas áreas del conocimiento. En este contexto, se presenta la segunda versión del clúster Cronos, una infraestructura basada en Raspberry Pi (ver [4]) diseñada para la enseñanza e investigación en computación paralela. Las mejoras implementadas incluyen la expansión del clúster de seis nodos a ocho, la incorporación de un nodo de login para la gestión de accesos, la sustitución del almacenamiento NFS basado en USB por un disco SSD y la integración de Shellinabox para permitir el acceso web. Estas optimizaciones tienen como objetivo mejorar la escalabilidad, la eficiencia y la administración del clúster, ofreciendo una experiencia de uso más robusta tanto para estudiantes como para investigadores.

Palabras Clave: Clúster, Raspberry Pi, cómputo paralelo, gestión de usuarios, escalabilidad.

CONTEXTO

El clúster Cronos fue desarrollado como una plataforma accesible para la enseñanza y experimentación en cómputo paralelo (ver [6]). Su diseño permite que los estudiantes trabajen en un entorno similar a infraestructuras HPC de mayor escala, brindando una base sólida para la comprensión de conceptos clave en procesamiento paralelo.

En diciembre de 2024, se llevó a cabo un workshop de introducción al cómputo paralelo utilizando Cronos (ver [6]). En este taller, más de cuarenta estudiantes de la carrera de programación adquirieron conocimientos sobre el diseño de algoritmos paralelos y realizaron prácticas en lenguaje C, utilizando herramientas y metodologías propias del cómputo paralelo. Durante las actividades, se identificaron limitaciones relacionadas con la administración de usuarios, la escalabilidad y el rendimiento del almacenamiento, lo que motivó la implementación de mejoras en la infraestructura del clúster.

A partir de este análisis, se decidió expandir el clúster de seis a ocho nodos para mejorar su capacidad de procesamiento, optimizar el almacenamiento compartido mediante un disco SSD y habilitar un nodo de login con doble interfaz de red, facilitando el acceso remoto mediante SSH y Shellinabox. Además, se planteó la futura implementación de OpenLDAP para la gestión centralizada de usuarios.

1. INTRODUCCIÓN

El procesamiento paralelo ha cobrado una relevancia significativa en la actualidad, permitiendo la ejecución eficiente de cálculos intensivos en diversas disciplinas. Los clústeres de computadoras proporcionan una infraestructura escalable y flexible para el desarrollo de aplicaciones científicas y académicas, pero su alto costo suele representar una barrera para instituciones educativas y centros de investigación con recursos limitados.

Las microcomputadoras de placa única (SBC), como la Raspberry Pi (ver [4]), han surgido

como una alternativa viable para la construcción de clústeres de bajo costo. Su bajo consumo energético y su arquitectura basada en procesadores ARM las convierten en una opción adecuada para la experimentación en computación distribuida, además de alinearse con los principios del Green Computing.

El concepto de Green Computing promueve la eficiencia energética y la reducción del impacto ambiental de los sistemas computacionales, encuentra un poderoso aliado. Al utilizar estas microcomputadoras en la construcción de clústeres, es posible crear soluciones de procesamiento paralelo que son no solo efectivas, sino también respetuosas con el medio ambiente.

El clúster Cronos fue concebido como una herramienta didáctica para la enseñanza de HPC y la experimentación en cómputo paralelo. Inicialmente compuesto por seis nodos, permitió a los estudiantes familiarizarse con tecnologías clave como Slurm (ver [2]), NFS y OpenMPI. Sin embargo, el uso creciente de la infraestructura evidenció la necesidad de ampliar su capacidad de cómputo, mejorar la administración de usuarios y optimizar el almacenamiento compartido. En este contexto, la evolución de Cronos busca ofrecer una plataforma más robusta y eficiente para la enseñanza y el desarrollo de proyectos de investigación en HPC.

El clúster Cronos utiliza el sistema operativo Linux Raspbian Bookworm (ver [5]), proporcionando un entorno estable y optimizado para la administración de tareas HPC.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El proyecto del clúster Cronos se desarrolla en torno a distintas líneas de investigación y desarrollo orientadas a mejorar la infraestructura de cómputo paralelo con hardware de bajo costo. Estas líneas de trabajo pueden dividirse en los siguientes ejes:

2.1 Expansión y escalabilidad del clúster

La incorporación de nuevos nodos busca analizar cómo el aumento de la capacidad de procesamiento impacta en el rendimiento del sistema y en la distribución de cargas de trabajo. Para ello, se han agregado dos nodos adicionales con hardware idéntico a los existentes. Se realizan pruebas de rendimiento utilizando herramientas como HPL Benchmark para evaluar la eficiencia del clúster tras la expansión y analizar la relación entre el número de nodos y el rendimiento general.

2.2 Optimización del almacenamiento compartido

Se investiga el impacto del cambio de un almacenamiento basado en USB a un disco SSD en el rendimiento de lectura/escritura y en la velocidad de acceso a datos por parte de los nodos. Se implementó un servidor NFS en el nodo maestro con el nuevo SSD como unidad de almacenamiento compartido. Se llevan a cabo pruebas con fio y dd para medir mejoras en la latencia y la velocidad de transferencia de archivos entre nodos.

2.3 Implementación de un nodo de login y gestión de accesos

El nodo de login permite mejorar la administración del clúster, ofreciendo un punto de acceso centralizado con doble interfaz de red, una conexión Ethernet dedicada a la red externa, por donde accederán los usuarios, y otra conexión interna que enlaza con la LAN del clúster. Esta separación de tráfico permite optimizar el rendimiento del clúster y garantizar la seguridad al aislar las comunicaciones internas de las externas. Además, en este nodo se configuró un servidor SSH para la autenticación y gestión de sesiones de trabajo, asegurando un acceso controlado y eficiente para los usuarios. Además, se instaló Shellinabox para

proporcionar acceso a una terminal desde un navegador web, facilitando la interacción con el clúster sin necesidad de un cliente SSH externo.

2.4 Administración centralizada de usuarios con OpenLDAP

La administración de usuarios en entornos HPC suele ser un desafío. Se investiga la implementación de OpenLDAP (ver [3]) para gestionar credenciales y permisos de manera eficiente en todos los nodos, con el objetivo de mejorar la seguridad y la facilidad de administración del clúster.

Cada una de estas líneas de investigación aporta conocimiento sobre el uso de clústeres de bajo costo para la educación y la experimentación en computación distribuida, alineándose con el propósito general del proyecto.

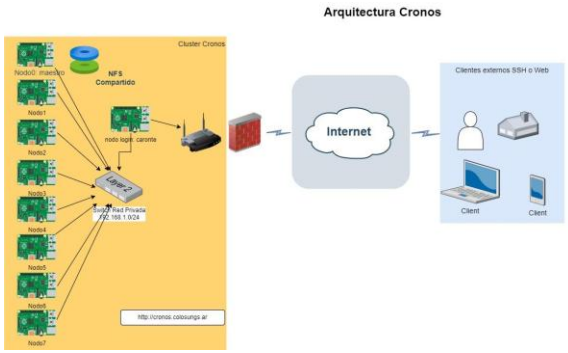


Figura 1 - Arquitectura de Cronos: nueva versión, ampliada y mejorada.

3. RESULTADOS OBTENIDOS/ESPERADOS

Hasta el momento, las modificaciones implementadas han generado mejoras en el rendimiento del clúster, específicamente en:

- **Acceso al almacenamiento compartido:** Se ha reducido el tiempo de acceso a datos en aproximadamente un 40% tras la sustitución del almacenamiento USB por un SSD. Para evaluar esta mejora, se realizaron pruebas con *fio* y *dd*, midiendo la latencia y la velocidad de transferencia de archivos antes y después del cambio. Los resultados mostraron un incremento significativo en la tasa de lectura/escritura, mejorando la eficiencia del sistema de archivos compartido.
- **Gestión de usuarios:** La incorporación del nodo de login ha permitido un acceso más controlado y seguro a la infraestructura.
- **Escalabilidad:**La adición de dos nodos ha incrementado la capacidad de procesamiento, habilitando la ejecución de tareas más complejas. Para evaluar el impacto de esta expansión, se midieron los tiempos de ejecución de procesos paralelos antes y después de la incorporación de los nuevos nodos, utilizando herramientas de benchmarking como time y GNU parallel. Además, se analizó la eficiencia del balanceo de carga mediante los registros de Slurm (ver [2]), verificando la distribución equitativa de tareas entre los nodos disponibles.
- **Accesibilidad:** La implementación de Shellinabox ha facilitado la interacción con el clúster sin requerir configuraciones adicionales en los clientes ya que pueden acceder a través de un navegador con conexión segura del protocolo https.

Como trabajos futuros, se planea:

- Implementar OpenLDAP (ver [3]) para la gestión centralizada de usuarios.
- Realizar pruebas de benchmarking para evaluar el rendimiento del clúster en diferentes escenarios.

- Explorar estrategias de optimización energética y balanceo de carga.
- Explorar fallas de hardware de un nodo y el plan de contingencia
- Copias de seguridad de datos y sistema operativo.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo está conformado por:

- Dr. Andrés Rojas Paredes – Director de investigación.
- Lic. Martha Semken – Investigadora.
- Lic. Mariano Vargas – Investigador.
- Ignacio Tula – Alumno adscripto de investigación.
- Giuliana Zorzoli – Alumna adscripta de investigación.

Este equipo permite una distribución eficiente de tareas dentro del proyecto, asegurando un equilibrio entre dirección, investigación y formación de nuevos profesionales en el área de HPC.

5. CONCLUSIONES

El clúster Cronos ha sido utilizado en diversas instancias de formación académica, incluyendo cursos y workshops sobre cómputo paralelo (ver [6]). La evolución de este cluster no solo beneficia a los estudiantes que utilizan la plataforma, sino que también brinda experiencia práctica en administración de sistemas, gestión de recursos y optimización de plataformas HPC. Asimismo, el trabajo de implementación de las líneas de investigación anteriormente descriptas, permitirá la capacitación de los investigadores, docentes y alumnos involucrados en el proyecto.

Además, este desarrollo contribuye a la democratización del acceso a tecnologías de cómputo paralelo en entornos educativos con recursos limitados, proporcionando una infraestructura accesible para la experimentación y formación en HPC.

A futuro, se plantea la posibilidad de explorar nuevas estrategias para optimizar el consumo energético del clúster y evaluar su integración con arquitecturas híbridas, permitiendo ampliar su aplicabilidad en distintos ámbitos de la investigación científica.

6. BIBLIOGRAFÍA

1. Pajankar, A. (2017). *Raspberry Pi Supercomputing and Scientific Programming*. Apress.
2. Slurm Workload Manager. (s.f.). *Slurm Documentation*. Recuperado de https://slurm.schedmd.com/documentation.html
3. OpenLDAP Project. (s.f.). *OpenLDAP Administrator's Guide*. Recuperado de https://www.openldap.org/doc/admin24/
4. The Raspberry Pi Foundation. (2023). *The Official Raspberry Pi Handbook*. Raspberry Pi Press.
5. Raspbian Project. (2023). *Raspbian Bookworm Documentation*. Recuperado de https://www.raspberrypi.com/documentation/computers/os.html
6. Semken, M., Vargas, M., Tula, I., & Rojas Paredes, A. (2024). *Configuración y puesta en marcha del clúster Raspberry Pi Cronos*. En XXX Congreso Argentino de Ciencias de la Computación (CACIC 2024), La Plata, Argentina. Recuperado de https://cacic2024.info.unlp.edu.ar/

CONTROL ADAPTADO PARA LA INTERACCIÓN NATURAL Y ACCESIBLE ROBÓTICA INCLUSIVA

Néstor Adrián Balich, Franco Adrián Balich,
Magali Bolivar, María Florencia Rossi, Matías Montiel, Romina Galizio

CAETI - Centro de Altos Estudios en Tecnología Informática
Universidad Abierta Interamericana. Informática (UAI)
Montes de Oca 745. Ciudad Autónoma de Buenos Aires, Argentina.

{nestor.balich, francoadrian.balich}@uai.edu.ar
{MagaliFlorencia.BolivarCruz, MariaFlorencia.Rossi, CarlaRomina.Galizio, MatiasNicolas.MontielTorres }@alumnos.uai.edu.ar

Resumen

El desarrollo de interfaces intuitivas y accesibles es un desafío clave en la evolución de la robótica, especialmente en contextos donde una interacción más natural con la tecnología mejora la experiencia del usuario. En este trabajo, presentamos un sistema de control innovador que permite operar un robot impreso en 3D mediante los movimientos del usuario, utilizando los sensores de movimiento (acelerómetro y giroscopio) de un dispositivo móvil.

Inicialmente concebido como una alternativa ergonómica para el control remoto, este enfoque demostró un gran potencial en accesibilidad, permitiendo que personas con diversas capacidades motrices interactúen con la tecnología de manera más intuitiva. La aplicación, desarrollada en App Inventor, se comunica con una placa ESP32 Nodemcu a través de Bluetooth 4.2 y controla un robot equipado con motores de corriente continua, tracción diferencial y un sensor de ultrasonido, permitiendo movilidad autónoma y teleoperación.

Más allá de su aplicación inmediata, este proyecto representa un primer paso hacia el desarrollo de nuevas estrategias inclusivas, explorando diferentes formas de interacción accesible con la tecnología. Además, en el ámbito educativo, facilita la enseñanza de robótica y programación en entornos interactivos, promoviendo el aprendizaje práctico e inclusivo en escuelas y centros de formación. [5]

Palabras clave: Móvil, Robot, ESP32, Dispositivos Móviles, IoT, MQTT, App Inventor, Acelerómetro, Giroscopio, Human Interface, HRI.

Contexto

El presente proyecto se desarrolla en el marco de las líneas de investigación del Laboratorio de Robótica Física e Inteligencia Artificial (LRFIA) de la Universidad Abierta Interamericana (UAI), dentro del Centro de Altos Estudios en Tecnología Informática (CAETI). Su enfoque principal es la aplicación de tecnologías emergentes en robótica y automatización educativa, promoviendo soluciones innovadoras que faciliten la interacción intuitiva con los sistemas automatizados.

En este contexto, el desarrollo de interfaces accesibles se presenta como un desafío clave para ampliar el alcance de la robótica, tanto en entornos educativos como en aplicaciones inclusivas. La posibilidad de controlar un robot mediante sensores de movimiento representa un avance significativo en la naturalidad de la interacción hombre-máquina, eliminando barreras de accesibilidad y facilitando el aprendizaje práctico.

Este proyecto cuenta con la participación de docentes y alumnos de las sedes Centro y Castelar (provincia de Buenos Aires), fomentando el trabajo colaborativo entre investigadores y estudiantes para la exploración de nuevas metodologías de enseñanza y el desarrollo de tecnologías con impacto social. [1], [3].

1. Introducción

El control de dispositivos robóticos ha evolucionado con la incorporación de tecnologías accesibles y herramientas de programación visual, facilitando su implementación en diversos entornos. En este contexto, App Inventor se presenta como una alternativa viable para el desarrollo rápido de aplicaciones de control, permitiendo la creación de interfaces intuitivas sin necesidad de conocimientos avanzados en programación.

Este proyecto propone el desarrollo de una aplicación móvil para operar un robot impreso en 3D, utilizando App Inventor, la IDE de Arduino y conectividad Bluetooth con una placa ESP32. La integración de sensores de movimiento, como el acelerómetro y el giroscopio, permite controlar el robot de manera natural e intuitiva, abriendo nuevas posibilidades de uso para personas con dificultades motoras o limitaciones en el manejo de interfaces convencionales.

Además de su potencial en accesibilidad, esta solución tiene un fuerte impacto en el ámbito educativo. La implementación de tecnologías accesibles en la enseñanza de la robótica permite reducir la curva de aprendizaje y fomentar un enfoque práctico basado en la experimentación. Promoviendo el desarrollo de habilidades en programación y electrónica, incentivando la creatividad y el pensamiento lógico en los estudiantes. [2], [4].

2. Línea de Investigación y Desarrollo

2.1 Conexión Bluetooth con ESP32

Para establecer la comunicación entre la aplicación móvil y el robot, se utilizó el módulo Bluetooth integrado en la ESP32. Esta conexión permite que la placa reciba los comandos enviados por la app y los traduzca en movimientos específicos, garantizando una interacción fluida y de baja latencia. El robot cuenta con dos motores de corriente continua manejados por un driver H que controla las ruedas de manera independiente. [1]

2.2 Uso de App Inventor

La aplicación móvil fue desarrollada en App Inventor, aprovechando su entorno de programación visual para facilitar su diseño e implementación. Se priorizó la creación de una interfaz intuitiva, minimizando el uso de texto e incorporando elementos gráficos de fácil comprensión, lo que permite su uso por una amplia variedad de usuarios sin necesidad de experiencia previa en programación. [4],[8].

2.3 Creación de la Interfaz UX

Con el objetivo de garantizar una experiencia accesible e intuitiva, se desarrolló una interfaz basada en iconografía visual y controles táctiles, eliminando la necesidad de habilidades

avanzadas de lectura o programación. En el contexto educativo, esta solución permite que los estudiantes interactúen con el robot sin barreras técnicas, promoviendo el aprendizaje basado en la experimentación.

La interfaz refleja en tiempo real los movimientos del usuario captados por el giroscopio y el acelerómetro del dispositivo móvil, permitiendo el control del robot mediante inclinaciones y gestos. Además, se incorporaron elementos visuales dinámicos, como indicadores de orientación y velocidad, que brindan retroalimentación instantánea sobre la dirección y respuesta del robot. Esta integración mejora la inmersión y la naturalidad en la interacción con el sistema.



Figura 1. Visualización de la interfaz de la aplicación de App Inventor.

2.4 Construcción del Robot

El chasis del robot se fabricó mediante impresión 3D, garantizando una estructura liviana y resistente. Se integró la ESP32 como microcontrolador principal y se optimizó el diseño para facilitar su ensamblaje y mantenimiento. [3],[6],[7].

2.5 Impacto en la Educación

La incorporación de esta tecnología en el ámbito educativo facilita el aprendizaje de robótica y programación al eliminar barreras técnicas y hacer que la experiencia sea más accesible para estudiantes de distintos niveles. Gracias a una interfaz intuitiva y visual, los alumnos pueden interactuar con el robot de manera sencilla, comprendiendo conceptos clave sin necesidad de conocimientos avanzados.

Este enfoque fomenta el aprendizaje basado en la experimentación y la resolución de problemas, promoviendo el pensamiento lógico y creativo. Además, su fácil implementación permite que docentes de diversas áreas puedan integrarlo en sus clases, ampliando las oportunidades de enseñanza en instituciones educativas. [5] [9] [10].

3. Resultados Obtenidos/Esperados

3.1 Resultados Obtenidos

- Se estableció con éxito la comunicación Bluetooth entre la ESP32 y la aplicación móvil.
- Se desarrolló una interfaz gráfica accesible en App Inventor, diseñada para facilitar su uso por cualquier persona. Gracias a la integración del acelerómetro y el giroscopio, el control del robot se realiza mediante movimientos intuitivos, eliminando la necesidad de botones complejos o conocimientos técnicos avanzados.
- Se construyó un robot funcional con chasis impreso en 3D.
- La interacción usuario-robot se logró de manera intuitiva, validando la accesibilidad e inclusión de la

aplicación, permitiendo que más personas puedan interactuar con la tecnología de manera sencilla y efectiva.

3.2 Resultados Esperados

- **Optimización del sistema de respuesta del robot,** mejorando la precisión y fluidez del control mediante ajustes en la interpretación de los datos del acelerómetro y giroscopio.
- **Incorporación de nuevas opciones de personalización,** permitiendo al usuario ajustar la sensibilidad del movimiento y adaptar la interfaz según sus necesidades.
- **Mejor estabilidad en la comunicación Bluetooth,** reduciendo posibles interferencias y asegurando una conexión más rápida y confiable entre la ESP32 y la aplicación móvil.
- **Ampliación de la retroalimentación visual en la interfaz,** agregando indicadores dinámicos que reflejan en tiempo real la dirección y velocidad del robot, mejorando la experiencia de usuario.

4. Formación de Recursos Humanos

El equipo de trabajo está conformado por:

- 2 docentes especialistas en IoT y robótica e impresión 3d, encargado de la supervisión técnica y metodológica del proyecto.
- 4 estudiantes de ingeniería en sistemas, responsables del desarrollo de la aplicación móvil, la integración de los sistemas electrónicos, la configuración y optimización del servidor MQTT y el desarrollo y optimización del chasis del robot mediante impresión 3D.

Bibliografia:

[1] R. A. Jaimes, "Control de Robot de Internet de las cosas (IoT) a través de la placa de desarrollo ESP32," Thesis for Telecommunications Engineering, June 2018. DOI: 10.13140/RG.2.2.27059.95522.

[2] "Smartphone Sensor-based Development and Implementation of a Remotely Controlled Robot Arm," JRC, vol. 5, no. 4, 2020. DOI: 10.18196/jrc.v5i4.21987.

[3] T. Lobato de Souza y L. Sato Elisiario, "Educational Robotics Teaching with Arduino and 3D Print Based on Stem Projects", Simposio Latinoamericano de Robótica (LARS) de 2019, Simposio Brasileño de Robótica (SBR) de 2019 y Taller sobre Robótica en Educación (WRE) de 2019 , Rio Grande, Brasil, 2019, pp. 407-410. DOI: 10.1109/LARS-SBR-WRE48964.2019.00078.

[4] "Using Scratch and App Inventor for teaching introductory programming in secondary education. A case study," IJTEL, 2016. DOI: 10.1504/IJTEL.2016.082317.

[5] M. C. Castañeda Barbarán, "Desarrollo de robots prototipos interactivos digitales para el aprendizaje de la programación con tecnología ARDUINO para fortalecer la enseñanza de las fases de robótica a docentes de UGEL Chiclayo," Bachelor's thesis, Universidad Nacional Pedro Ruiz Gallo, 2018. DOI: 20.500.12893/9178

[6] G. López Valero, "Diseño, estudio y simulación de un robot bípedo en

CoppeliaSim (V-REP) y construcción de un prototipo en impresión 3D," Bachelor's thesis, Universitat Politècnica de València, 2020. Recuperado de: https://riunet.upv.es/handle/10251/153249

[7] Gomez, J. & Macias, J. C. (2022). Diseño e Implementación de un Robot Paralelo 2RR para Aplicaciones de Impresión 3D.. Recuperado de: http://hdl.handle.net/20.500.11912/10968.

[8] Poveda-Pineda, Derly F., & Cifuentes-Medina, José E.. (2020). Incorporación de las tecnologías de información y comunicación (TIC) durante el proceso de aprendizaje en la educación superior. Formación universitaria, 13(6), 95-104. DOI: 10.4067/S0718-50062020000600095

[9] M. Kandlhofer, G. Steinbauer, S. Hirschmugl-Gaisch and P. Huber, "Artificial intelligence and computer science in education: From kindergarten to university," 2016 IEEE Frontiers in Education Conference (FIE), Erie, PA, USA, 2016, pp. 1-9, DOI: 10.1109/FIE.2016.7757570.

[10] Herrera Pérez, Y. J., & Espinosa Cevallos, P. A. (2024). Impacto de la enseñanza basada en proyectos apoyada por tecnología en el desarrollo de habilidades del siglo XXI en estudiantes de secundaria. Bastcorp International Journal, 3(1), 4–18. DOI: 10.62943/bij.v3n1.2024.33.

Del aula invertida al modelo tecnopedagógico en la enseñanza del paradigma paralelo

Caymes-Scutari Paola[1,2,3], Bianchini Germán[1], Espinosa Morata M. Tomás[3]

[1]Laboratorio de Investigación en Cómputo Paralelo/Distribuido,
Departamento de Ingeniería en Sistemas de Información. Facultad Regional Mendoza/Universidad Tecnológica Nacional
Rodríguez 273 (M5502AJE) Mendoza, +54 261 5244579
[2]Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)
[3]Facultad de Ingeniería. Universidad Nacional de Cuyo, Centro Universitario (M5502JMA) Mendoza.

pcaymesscutari@frm.utn.edu.ar, gbianchini@frm.utn.edu.ar,
mariano.espinosa@ingenieria.uncuyo.edu.ar

RESUMEN

La educación mediada por tecnología se encuentra en un momento de pleno desarrollo, facilitada por la existencia de entornos virtuales de enseñanza aprendizaje, el acceso y uso creciente de distintos tipos de dispositivos (pc, tablet, celular, etc.), y las facilidades de conectividad. La misma posibilita la evolución del proceso de enseñanza-aprendizaje más allá del tiempo y el espacio establecido generalmente por las clases presenciales, y da lugar a otras formas de interacción, de aprender y de regular el autoaprendizaje. A su vez, en los últimos años se ha venido suscitando y fomentando por parte de las instituciones un cambio de paradigma educativo, centrado en el estudiante, por lo que las metodologías activas resultan primordiales al pensar un espacio curricular. En este contexto, presentamos una propuesta que parte de la experiencia adquirida en base a un modelo subyacente de aula invertida, el cual se ha rediseñado y enriquecido para el reemplazo de horas presenciales por horas virtuales, para la enseñanza del paradigma paralelo, en el área de informática. El artículo describe los rasgos principales de la propuesta de base y los resultados, como así también los elementos que enriquecen la nueva propuesta para el ciclo 2025.

***Palabras clave*:** EVEA, entorno virtual de enseñanza-aprendizaje, metodologías activas, evaluación contínua, seguimiento.

CONTEXTO

La presente propuesta tuvo su inicio en la asignatura "Computación Paralela" (asignatura electiva de tercer año, de la carrera de Ingeniería en Sistemas de Información de la UTN Facultad Regional Mendoza) para el desarrollo de la cual una de las metodologías utilizadas la constituye el modelo de aula invertida (Martínez-Olvera, Esquivel-Gámez y Martínez Castillo, 2014). La misma ha sido implementada desde el ciclo 2020, momento en el que la pandemia de Covid propició y requirió la implementación de un modelo que ya estaba en estudio y

preparación. Posteriormente, el mismo modelo se reprodujo desde el ciclo 2023 en la asignatura "Programación Paralela y Distribuida" (asignatura obligatoria de cuarto año, de la carrera de Licenciatura en Ciencias de la Computación de la Universidad Nacional de Cuyo) en la que se ha aplicado de forma similar. En esta última institución, en el marco de un proyecto institucional para implementar espacios curriculares con reemplazo de horas presenciales por horas a distancia, para el ciclo 2025 se prevé enriquecer el modelo con la implementación de un modelo tecno-pedagógico (Castellanos Santos, Sánchez Romero y Calderero Hernández, 2017) que contempla precisamente la realización de actividades a distancia. En este contexto, la experiencia adquirida previamente con el modelo de aula invertida y las actividades de seguimiento, constituyen una base importante para poder dar el siguiente paso, en una propuesta que conjugue y equilibre presencialidad y asincronismo.

1. INTRODUCCIÓN

La metodología de enseñanza-aprendizaje que llevamos adelante desde 2020 considera distintos elementos, instancias y estrategias, tales como clases, trabajos prácticos, trabajos de laboratorio y un trabajo global integrador, a fin de promover el desarrollo integral de los resultados de aprendizaje propuestos. En este esquema, en general, se propone adoptar una metodología con clases teórico-prácticas, con características de aula invertida (Martínez-Olvera, Esquivel-Gámez y Martínez Castillo, 2014), que contemple los momentos de autoaprendizaje, co-aprendizaje y socialización, como así también integrando el proceso de evaluación contínuo, tanto formativo como sumativo (Casanova, 1998). De acuerdo a dicho esquema, se asigna con antelación el material que los estudiantes deben revisar de forma autónoma antes del siguiente encuentro (videos, bibliografía, etc.). Durante las clases se realiza una puesta en común, mediada a través de diferentes recursos (cuestionarios, encuestas, juegos, infografías, etc.) que permita identificar los conceptos o aspectos que requieren un tratamiento adicional. En esta fase de co-aprendizaje y co-evaluación formativa el docente responsable regula en qué medida es necesario reformular los temas que conforman el programa utilizando como apoyo para el mismo el uso de la pizarra, la presentación de diapositivas con proyector multimedia, o la demostración de código, lo cual brinda la posibilidad de reforzar explicaciones y mejorar la comunicación docente-estudiante, y cada estudiante es responsable de solicitar la asistencia necesaria. A su vez, se propicia e impulsa la discusión de los temas en clase, ganando mayor reflexión, interacción e interés por parte del estudiante y la puesta en común. En las clases, momentos o asignaciones de carácter más práctico o de laboratorio, se tienen en cuenta las necesidades a atender en cuanto a los algoritmos y/o programas que se estén desarrollando.

A partir de la síntesis anterior, es posible apreciar que el enfoque adoptado está centrado en el estudiante (McCombs, 2000) a través de la implementación de metodologías activas (Labrador Piquer y Andreu Andrés, 2008) como aprendizaje invertido, la integración de la evaluación durante todo el proceso de enseñanza-aprendizaje, el uso de encuestas (Moodle Encuestas, 2025) o cuestionarios para el seguimiento formativo clase a clase, el uso de cuestionarios kahoot! (Kahoot, 2025) para repasos lúdicos previos a las instancias de evaluación sumativa (Casanova, 1998), y el desarrollo de

habilidades sociales y comunicativas a través del trabajo en equipo y comunicación de los resultados alcanzados. Con todo ello se intenta poner en juego el "hacer" y el "resolver" dentro del aula, con el acompañamiento de los docentes, y así lograr aprendizajes y/o producciones más significativos y razonados. Todas estas actividades confluyen en el aula abierta de la asignatura, implementada sobre la plataforma Moodle (Moodle, 2025), la cual se utiliza para poner a disposición el material de estudio y referencia, intercambiar material o información, como así también para acceder a foros, atender consultas, realizar actividades, encuestas, etc.

2. LÍNEAS DE INVESTIGACIÓN y DESARROLLO

Para el ciclo 2025 se propone implementar un modelo tecno-pedagógico (Castellanos Santos, Sánchez Romero y Calderero Hernández, 2017) de virtualización de espacios curriculares, motivo por el cual la propuesta educativa debe tener en cuenta las características de la asignatura, de los estudiantes, y del entorno virtual de enseñanza-aprendizaje, a fin de seleccionar las metodologías, recursos, y estrategias que permitan repartir de forma equitativa la carga horaria dedicada a actividades presenciales y la carga horaria dedicada a actividades a distancia y/o asíncronas. La propuesta, en general, comparte los elementos y recursos descriptos en la sección 1 (clases, trabajos prácticos, trabajo global integrador, encuestas, foros, kahoot!, etc.), sobre las que incorpora o rediseña algunas actividades guiadas –actividades de apertura de cada unidad temática (Tobón Tobón, Pimienta Prieto y García Fraile, 2010), y uso de documentos de Colab (Colab, 2025)–, así como también contempla una integración más fuerte del proceso de autoevaluación para las horas dedicadas al aprendizaje autónomo y de coevaluación para las horas dedicadas al aprendizaje presencial en el aula. Todo ello orientado a brindar las herramientas necesarias para facilitar el autoaprendizaje y a valorar la capacidad de autovaloración y autocrítica que todo futuro profesional debería tener en el ejercicio de la profesión (Kowalski et al, 2019).

El aula virtual ha sido rediseñada para que el trayecto sea más guiado y autocontenido, en el sentido de que hemos incorporado explicaciones, reflexiones, consignas, etc., a fin de que el recorrido por las actividades y los temas tenga un hilo conductor aún sin la presencia de los docentes. En los párrafos que siguen, comentamos en más detalles algunos de los elementos que incorpora la nueva propuesta.

Las actividades de apertura de cada Unidad temática se proponen para su desarrollo de forma asíncrona antes del abordaje del material de estudio asignado para la clase invertida, y se retoman posteriormente de forma conjunta durante la clase presencial. El objetivo de la actividad de apertura consiste en ofrecer elementos disparadores para la reflexión, a partir de aspectos conocidos de la vida cotidiana, que tengan un punto de encuentro o conexión con el objeto de estudio a desarrollar en la unidad temática, y que faciliten la introducción a dicho tema. Dado que se trata de una actividad de apertura y de opinión, es importante que cada estudiante pueda manifestar su punto de vista libremente, brindando su propia interpretación de la propuesta, sin que su aporte se vea influenciado o interpelado por las consideraciones de sus pares. Pero al mismo tiempo, también es muy valiosa la

construcción colectiva que se realiza a partir del análisis de las coincidencias y diferencias entre los aportes de cada estudiante. En este contexto, hemos propuesto que cada estudiante participe en un foro de tipo PyR (Moodle Foros, 2025), plasmando la reflexión originada a partir de la actividad de apertura. En este tipo de foros, cada estudiante sólo puede visualizar las respuestas de sus pares una vez que haya enviado su propio aporte. De esta manera, se logra el triple objetivo de tener una evidencia de la realización de la actividad, de conocer la reflexión personal de cada estudiante, y que a su vez cada quien pueda acceder a las variadas opiniones del resto de pares para llegar a una reflexión colectiva posiblemente más completa y superadora.
En lo que respecta a los antes mencionados documentos de Colab, se propone incorporarlos para asistir el desarrollo asíncrono de las actividades prácticas de programación, ya que ofrecen la posibilidad de brindar un entorno unificado, con ejemplos de uso y ejecución, sobre los que cada estudiante puede experimentar, así como también desarrollar sus propios programas.

3. RESULTADOS OBTENIDOS

Los resultados obtenidos desde 2020 con la modalidad presentada en la sección 1 han sido muy positivos, tanto tomando en consideración los resultados estadísticos en cuanto al porcentaje de estudiantes que regularizaron y que promocionaron la asignatura, como la opinión de los propios estudiantes expresada directamente o a través de la encuesta final del cursado, valorando positivamente el cambio de foco y la incorporación de las actividades de seguimiento y lúdicas. En cuanto al nuevo modelo con reemplazo de horas presenciales por horas a distancia, esperamos que la experiencia también sea positiva, que permita flexibilizar la dedicación y organización horaria de los estudiantes para congeniar lo académico con lo laboral, y sobre todo que resulte satisfactoria para los estudiantes que asuman su protagonismo con responsabilidad.

4. FORMACIÓN DE RECURSOS HUMANOS

El aporte en cuanto a formación de recursos humanos, puede considerarse en cuanto a la formación de cada estudiante, como de cada docente.

En relación a los estudiantes, la presente propuesta tiene como objetivo innovar en cuanto a la forma de implementar el proceso de enseñanza-aprendizaje para esta asignatura, brindando la posibilidad de que los estudiantes tengan acceso a otras formas de aprender, combinando lo ubicuo con lo presencial, e integrando las actividades evaluativas de seguimiento a cada paso, formándolos y acompañándolos para que adquieran experiencia en el aprendizaje o formación autónomos (Kowalski et al, 2019).

En cuanto a los docentes, tanto la implementación del modelo basado en aula invertida como el nuevo modelo tecno-pedagógico propuesto han requerido la capacitación de los docentes de ambas asignaturas a través de cursos, seminarios, y revisión del estado del arte, tanto en lo que hace a la formación por competencias (Tobón Tobón, 2013) como al aprendizaje centrado en el estudiante, (McCombs, 2000) y la educación

mediada por entornos virtuales (Arango-Vázquez y Marique-Losada, 2023).

5. BIBLIOGRAFÍA

Arango-Vázquez, S. y Marique-Losada, B. (2023). Interacciones comunicativas y colaboración mediada por entornos virtuales de aprendizaje universitarios. Revista de Educación a Distancia. Universidad de Murcia. Vol. 23 Núm. 76 (2023): Ordinario

Casanova, M. A. (1998). La evaluación educativa. México: Muralla.

Castellanos Santos, A., Sánchez Romero C. y Calderero Hernández, F. (2017). Nuevos modelos tecnopedagógicos. Competencia digital de los alumnos universitarios. Revista Electrónica de Investigación Educativa. REDIE vol.19 no.1. https://doi.org/10.24320/redie.2017.19.1.1148

Colab – Google Colaboratory (2025). Accedido en febrero 2025. https://colab.google/

Kahoot! (2025). Accedido en febrero 2025. https://kahoot.com/

Kowalski, V., Morano, D., Erck, I., Cirimelo, S., y Enríquez H. (2019). ¿Qué debemos cambiar para orientarnos a un enfoque basado en competencias?. Facultad de Ingeniería, Universidad Nacional de Misiones.

Labrador Piquer, M. J., y Andreu Andrés, M. Á. (2008). Metodologías Activas. Valencia: Editorial de la UPV.

Martínez-Olvera, W., Esquivel-Gámez, I. y Martínez Castillo J. (2014). Los Modelos Tecno-Educativos, revolucionando el aprendizaje del siglo XXI (pp.143-160) Edition: PrimeraChapter: Aula Invertida o Modelo Invertido de Aprendizaje: origen, sustento e implicaciones.DSAE-Universidad Veracruzana.

McCombs, B. (2000). Assessing the Role of Educational Technology in the Teaching and Learning Process: A Learner-Centered Perspective. Accedido en febrero 2025, de ERIC: https://files.eric.ed.gov/fulltext/ED452830.pdf

Moodle (2025). Accedido en febrero 2025. https://moodle.org/?lang=es

Moodle Encuestas (2025). Módulo de Encuestas. Accedido en febrero de 2025. https://docs.moodle.org/all/es/M%C3%B3dulo_de_encuesta_predefinida

Moodle Foros (2025). Actividad de Foro. Accedido en febrero de 2025. https://docs.moodle.org/all/es/Actividad_de_foro

Tobón Tobón, S. (2013). Formación basada en competencias: pensamiento complejo, diseño curricular y didáctica (Segunda ed.). Bogotá: Ecoe Ediciones.

Tobón Tobón, S., Pimienta Prieto, J., García Fraile, J. (2010). Secuencias Didácticas: Aprendizaje y Evaluación de Competencias. México: Pearson Educación.

EL IMPACTO DE LA INTELIGENCIA ARTIFICIAL EN LA DEFINICIÓN, DESARROLLO Y EVALUACIÓN DE COMPETENCIAS

Carlos Neil, Nicolás Battaglia, Marcelo De Vincenzi,
Nelson Garrido, Fabian Maffei, Silvia Poncio, Marcelo Monferrato

Universidad Abierta Interamericana. Facultad de Tecnología Informática.
Centro de Altos Estudios en Tecnología Informática. Buenos Aires. Argentina
{carlos.neil, nicolas.battaglia, medevincenzi}@uai.edu.ar
{nelson.garrido, fabian.maffei, silviavictoria.poncio, marcelo.monferrato}@alumnos.uai.edu.ar

RESUMEN

La incorporación de la Inteligencia Artificial (IA) está transformando la educación y, en este contexto, el aprendizaje como reorganización de las estructuras cognitivas impele distinguir entre el *proceso* y el *producto* del aprendizaje. El primero, abarca las actividades mentales para lograr un aprendizaje significativo; el *producto*, se visualiza en los resultados observables producidos por los estudiantes. Previo a las Tecnologías de la Información y las Comunicaciones (TIC) se asumía una correlación directa entre los productos observables y el aprendizaje del estudiante; aquel era un reflejo de éste. La irrupción de las TIC, potenciadas por la IA, ha desafiado esa lógica debido a que la generación automática de *productos* reduce su valor como indicador del aprendizaje. A partir de estos conceptos, el Aprendizaje Centrado en el Estudiante (ACE) encuentra un obstáculo serio con el uso indiscriminado de la IA debido a las limitaciones que implica su uso en el desarrollo de las habilidades cognitivas. En el otro extremo de la relación docente/estudiante, el uso de la IA abre un abanico amplio de posibilidades para los docentes y emerge como un aliado en el diseño y desarrollo de las actividades que requieren mucho esfuerzo y tiempo.

Palabras clave: inteligencia artificial; aprendizaje centrado en el estudiante; desarrollo de competencias; evaluación.

CONTEXTO

El proyecto de investigación se desarrolla en la Facultad de Tecnología Informática de la UAI, donde se implementan las estrategias didácticas y herramientas derivadas de las investigaciones en asignaturas de grado y posgrado.

1. INTRODUCCIÓN

Este trabajo es una continuación del presentado en [1], focalizado en la definición, desarrollo y evaluación de competencias en carreras tecnológicas, pero ahora enmarcado en las implicancias de la utilización de la IA.

En lo que continúa, se reflexiona sobre el uso de la IA en la educación y en la necesidad de diferenciar su uso por estudiantes y docentes. Luego, se propone al enfoque constructivista como la base teórica que sustenta tanto el ACE como el trabajo colaborativo y es, en este aspecto, donde se considera el impacto negativo IA. Después, se presenta el concepto de ACE como faro que guía toda la propuesta educativa. Por otro lado, se enfatiza en la importancia del trabajo en equipo y la colaboración. Presentados los fundamentos, se plantea el modelo por competencias como una forma de implementar el ACE y el uso de rúbricas en la evaluación y los lineamientos, derivados de aquellas, como guías en el proceso de aprendizaje. Como concepto integrador, la definición, el desarrollo y la evaluación de competencias se estructuran considerando el alineamiento constructivo, que enfatiza la necesaria coherencia entre estos tres componentes.

1.1 Inteligencia artificial en la educación

Los actuales avances en la IA están generando transformaciones profundas en el ámbito educativo, redefiniendo paradigmas de aprendizaje y enseñanza [2]. Estas tecnologías no solo ofrecen acceso a una cantidad sin precedentes de información, sino que también introducen herramientas innovadoras que transforman la manera en que la concebimos [3] creando experiencias adaptativas que no estaban disponibles en los ámbitos educativos previos [4]. Por esta razón, la incorporación de la IA en la educación nos impela a reconsiderar los métodos tradicionales tanto en la enseñanza como en la evaluación que resultan inadecuados frente a las nuevas dinámicas educativas.

Si concebimos el aprendizaje como una transformación de las estructuras cognitivas [5], es importante distinguir entre su *proceso* y el *producto* resultante. El primero, entendido como

las actividades mentales que el estudiante desarrolla, constituyen la base del aprendizaje significativo [6]. El *producto*, por otro lado, se manifiesta visible en los comportamientos observables de los estudiantes. Si no se distinguen ambos, existe el riesgo de centrarse exclusivamente en los resultados finales (*productos*) y descuidar los procesos de análisis y las estrategias cognitivas y metacognitivas (*procesos*) que condujeron a esos resultados

Resaltamos que, aunque ambos conceptos están vinculados, el *proceso* y el *producto* son dimensiones diferentes en el proceso de enseñanza/aprendizaje. La transformación de las estructuras cognitivas depende de un proceso activo y significativo en donde los docentes deben utilizar estrategias de aprendizaje efectivas (*proceso*) y no solo evaluar los resultados (*producto*).

Previo a la aparición de las TIC se asumía que el *producto* del estudiante era un reflejo fiel de su *proceso* de aprendizaje, estableciéndose una correlación directa y casi automática entre ambos. Sin embargo, la irrupción de las TIC, potenciadas por la IA, ha cuestionado esa lógica tradicional. La generación automática de *productos* reduce (o anula) la percepción de su valor como indicador del aprendizaje. En este contexto, el proceso de aprendizaje queda en un segundo plano, ya que, al no ser visible ni evaluable, tiende a desestimarse, quedando el *producto* como resultado engañoso del proceso de aprendizaje. Esta realidad impone un condicionante importante en las propuestas educativas que los docentes deberán utilizar para desarrollar las competencias esperadas.

1.2 Inteligencia artificial: el uso en estudiantes y docentes

La IA ha emergido como una herramienta clave en la educación, ofreciendo soluciones personalizadas y optimizando procesos de enseñanza/aprendizaje. Sin embargo, el uso de la IA debe diferenciarse según el rol que desempeñan los actores principales: docentes y estudiantes. El proceso de aprendizaje se construye mediante la participación del estudiante y el inevitable uso de la IA plantea desafíos en relación con el equilibrio entre facilitar el aprendizaje y la adquisición de resultados educativos sin esfuerzo. Si bien la IA puede ser un aliado poderoso, también puede constituirse en una amenaza si se la utiliza de manera inapropiada. Las herramientas de IA pueden aliviar la carga cognitiva del estudiante, pero también se corre el riesgo de limitar su participación en el proceso de aprendizaje, sin obviar los temas éticos vinculados [7]. El uso indiscriminado de la IA puede fomentar un aprendizaje *pasivo* en el que los estudiantes se convierten en consumidores de respuestas o soluciones en lugar de productores de conocimiento y que, en el mejor de los casos, se transformen en competentes escritores de *Prompts*. Para abordar estos desafíos, es crucial que la IA se utilice como una herramienta complementaria al proceso de aprendizaje, no como un sustituto del esfuerzo estudiantil. Esto implica que los docentes deben diseñar actividades que enfaticen el *proceso* más que el *producto* final.

Por otro lado, en las múltiples actividades docentes, que implican mucho esfuerzo y tiempo en su diseño y desarrollo, el uso de la IA sí abre un abanico amplio de posibilidades que van desde la evaluación eficiente y objetiva mediante la automatización de procesos de corrección, pasando por el diseño y la personalización curricular y llegando al desarrollo de sistemas de gestión del aprendizaje que ayuden a identificar lagunas de conocimiento y propongan estrategias didácticas.

1.3. Enfoque constructivista

Desde la perspectiva constructivista el aprendizaje es un proceso activo donde los estudiantes construyen su conocimiento a través de experiencias contextualizadas. Fue propuesto inicialmente por Piaget [8] a partir de la teoría del desarrollo cognitivo, posteriormente expandida por Vygotsky [9]. Por otro lado, Ausubel [6], enfatizó en el aprendizaje significativo, donde los nuevos conocimientos se relacionan con los previos del estudiante, utilizando organizadores y materiales que faciliten la asimilación de la información y su vínculo con otros conceptos.

A pesar de las dudas planteadas sobre el uso de la IA, ésta puede potenciar el enfoque constructivista al proporcionar experiencias de aprendizaje personalizadas y contextualizadas a partir del análisis de los conocimientos previos del estudiante proponiendo materiales y actividades específicas que conecten con ellos, facilitando el aprendizaje significativo. Además, puede generar entornos interactivos y simulaciones que permitan experimentar y construir conocimiento de manera activa.

Otras perspectivas también interactúan con la IA. El conductismo, por ejemplo, a través de herramientas, como los sistemas de aprendizaje adaptativo, que utilizan la IA para ofrecer refuerzos inmediatos y personalizados [4]. Además, el conectivismo, con su enfoque en el aprendizaje a través de redes y conexiones,

encuentra en la IA una poderosa herramienta para organizar información y fomentar interacciones en entornos digitales [10].

1.4. Aprendizaje centrado en el estudiante

El ACE es un enfoque central en la educación [11], se basa en teorías constructivistas que proponen que los estudiantes deben construir y reconstruir el conocimiento para aprender de forma efectiva. Es también un proceso de cambio cualitativo, reforzando y empoderando al estudiante, desarrollando sus habilidades críticas [12]. La IA puede favorecer el ACE al ofrecer herramientas que promuevan la construcción activa y autónoma del conocimiento a través de plataformas que puede adaptarse al ritmo y estilo de aprendizaje de cada estudiante, además de ofrecer herramientas que personalizan y optimizan el aprendizaje facilitando la retroalimentación en tiempo real, permitiendo a los estudiantes identificar y abordar sus áreas de dificultad de manera inmediata [13].

1.5. Trabajo colaborativo

El trabajo colaborativo, fundamentado en el concepto de zona de desarrollo próximo [14] ha experimentado un notable impulso. Una forma de implementarlo es mediante el aprendizaje basado en problemas que fomenta el pensamiento complejo y la reflexión grupal y cooperativa para tomar decisiones sobre problemas reales y relevantes dentro del ámbito profesional en el que participa y se forma el estudiante [15]. La IA puede potenciar el trabajo colaborativo y el aprendizaje basado en problemas al crear entornos virtuales donde los estudiantes colaboren para resolver problemas reales. Asistentes inteligentes pueden facilitar la comunicación y coordinación entre los miembros del grupo, mientras que algoritmos de IA pueden generar problemas adaptados al nivel de los estudiantes, promoviendo el pensamiento crítico y complejo.

1.6. Competencias y resultados de aprendizaje

El concepto de competencia es amplio; algunos la definen en términos de rendimiento y habilidades adquiridas a través de la formación, otros adoptan una perspectiva más holística que engloba el conocimiento, la comprensión, las habilidades, las destrezas y las actitudes. Los resultados de aprendizaje, más específicos, son formulaciones que indican lo que se espera que los estudiantes conozcan, sean capaces de hacer o demostrar al finalizar un período formativo [16]. Este proyecto se enfoca en el modelo por competencias como una de las implementaciones del ACE. La matriz de competencias, por otro lado, ofrece una visión general de la distribución de responsabilidades en el desarrollo y evaluación de competencias en los espacios curriculares considerando los diferentes niveles de dominio del estudiante [17].

La IA puede contribuir al modelo por competencias al facilitar tanto el desarrollo como la evaluación de manera personalizada. Sistemas basados en IA pueden considerar a las habilidades, conocimientos y actitudes de los estudiantes en tiempo real, alineándolos con los resultados de aprendizaje esperados mediante el análisis de la matriz de competencias, identificando áreas de mejora y distribuyendo las responsabilidades de manera más eficiente en los espacios curriculares.

1.7. Evaluación y lineamientos para el aprendizaje utilizando rúbricas

Las rúbricas brindan criterios de calidad relacionados con las competencias y son un instrumento válido para su evaluación [18]. La autoevaluación basada en su uso incorpora los criterios que los estudiantes deben aplicar para valorar su propio trabajo [19]. Las rúbricas, por otro lado, no se limitan únicamente a la evaluación, también tienen un componente instructivo transformándose en una estrategia de aprendizaje donde el estudiante no resuelve un problema y luego se autoevalúa, sino que lo desarrolla con las consideraciones establecidas en los criterios de evaluación, devenidos en lineamientos para el aprendizaje, que le permite reflexionar y actuar en consecuencia [20]. Estas tecnologías pueden enriquecer el uso de rúbricas al automatizar la retroalimentación basada en los criterios de evaluación, ayudando a los estudiantes a reflexionar y mejorar en tiempo real. Además, la IA puede utilizar las rúbricas como herramienta instructiva, adaptando las actividades de aprendizaje a las necesidades del estudiante.

1.8. Alineamiento constructivo

El alineamiento constructivo es un enfoque pedagógico reconocido por su énfasis en la necesaria coherencia que debe existir entre los resultados de aprendizaje, el proceso educativo y la evaluación [21]. En esta línea, es fundamental que los resultados de aprendizaje, las actividades de enseñanza, así como los métodos de evaluación se definan de manera coherente para

maximizar el aprendizaje. La IA puede coadyuvar a alineamiento constructivo mediante la sugerencia de actividades y evaluaciones que estén directamente vinculadas con los resultados de aprendizaje esperados, asegurando que cada uno de los componentes del proceso educativo estén alineado entre sí.

.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El proyecto de investigación se desarrolla en el Centro de Altos Estudios en Tecnología Informática (CAETI) de la UAI y se enmarca en la línea de investigación *Sociedad del conocimiento y tecnologías aplicadas a la educación*.

3. RESULTADOS OBTENIDOS / ESPERADOS

3.1. Resultados obtenidos

En el marco del proyecto de investigación se han presentado, en 2024, los siguientes artículos científicos: *Definición, desarrollo y evaluación de competencias en carreras de informática* [1]; *La inteligencia artificial en la educación y sus implicaciones: un mapeo sistemático de la literatura* [22]; *Teorías de aprendizaje en el contexto de la Inteligencia Artificial* [23]; *Rubric-Driven Competency Development: A Case Study* [24]. Además, se editó el libro *Marco conceptual para la definición, desarrollo y evaluación de competencias* [25] y se desarrolló el *Taller de definición, desarrollo y evaluación de competencias* con la participación de docentes de Argentina y Colombia en paralelo con tres webinar para docentes de carreras tecnológicas [26].

3.2. Resultados esperados

En 2025, se desarrollarán trabajos que se verán reflejados en artículos científicos, tesis y trabajos finales de posgrado.

A partir de la aplicación del proceso de desarrollo conducido por rúbricas [24]: 1) el diseño de mapas conceptuales (*Rubric-Driven Concept Maps Design*); 2) la escritura de resultados de aprendizaje (*Rubric-Driven Learning Outcomes Writing*) que amplía [27] y, 3) el diseño de rúbricas (*Rubric-Driven Rubric Design*) que amplía [28]. Como parte de tesis y trabajos finales de posgrado, se desarrollarán: 4) el impacto de la IA en el aprendizaje (*Impact of Artificial Intelligence on the Learning Experience*: *A Systematic Literature Mapping*); 5) la Automatización en la generación de resultados de aprendizaje *(Automation of Learning Outcome Generation through Large Language Models and Augmented Retrieval*); 6) *Percepción de estilos de aprendizaje y personalización con Inteligencia Artificial*. Como implementaciones tecnológicas: 7) el rediseño de UAIRubric (case.uai.edu.ar/rubrics/) mediante un framework para aplicaciones Web, utilizando bases de datos SQL y noSQL mediante microservicios y 8) el análisis de estilos de aprendizaje utilizando herramientas de IA. Por último, se diseñará un marco conceptual para el desarrollo de competencias utilizando micro credenciales.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de investigación está compuesto por estudiantes de grado y posgrado de la UAI, que están desarrollando sus respectivas tesis y trabajos finales en temas relacionados con el proyecto: Nelson Garrido, Fabian Maffei, Julian Escobedo, Silvia Poncio, Gabriela Iannantuoni, Charles Maldonado, Marcelo Monferrato, Mauricio Hernandez y Facundo Romeu.

5. BIBLIOGRAFÍA

[1] M. Neil, C. G., Battaglia, N., De Vincenzi, M, Maffei, F., Garrido N., Poncio, S., Oliva, F., Lomoro, J., Iannantuoni, G., Maldonado C., Monferrato, "Definición, desarrollo y evaluación de competencias en carreras de informática," in *XXVI Workshop de Investigadores en Ciencias de la Computación (WICC)*, 2024.

[2] M. M. O. Fernández and A. A. Ferreiro, "Las tecnologías digitales en el entrenamiento de las funciones ejecutivas: una revisión sistemática de literatura," *RiiTE Rev. Interuniv. Investig. en Tecnol. Educ.*, pp. 120–136, 2023.

[3] C. Beaulac and J. S. Rosenthal, "Predicting university students' academic success and major using random forests," *Res. High. Educ.*, vol. 60, pp. 1048–1064, 2019.

[4] H. Peng, S. Ma, and J. M. Spector, "Personalized adaptive learning: an emerging pedagogical approach enabled by a smart learning environment," *Smart Learn. Environ.*, vol. 6, no. 1, pp. 1–14, 2019.

[5] D. B. Novak, J. D. y Gowin, *Aprendiendo a Aprender*. Martínez Roca, 1988.

[6] D. Ausubel, J. Novak, and H. Hanesian, "Significado y aprendizaje significativo,"

Psicol. Educ. un punto vista cognoscitivo, vol. 1, no. 2, pp. 53–106, 1976.
[7] A. Nguyen, H. N. Ngo, Y. Hong, B. Dang, and B.-P. T. Nguyen, "Ethical principles for artificial intelligence in education," *Educ. Inf. Technol.*, vol. 28, no. 4, pp. 4221–4241, 2023,
[8] L. S. Vygotsky, *The collected works of LS Vygotsky: Problems of the theory and history of psychology*, vol. 3. Springer Science & Business Media, 1987.
[9] J. Piaget, M. Cook, and W. W. Norton, "The origins of intelligence in children, vol. 8 New York," *NY Int. Univ. Press. Sch.*, 1952.
[10] G. Siemens, "Connectivism: A learning theory for the digital age. International Journal of Instructional Technology and Distance Learning," *Online] retrieved from http//www. idtl. org/Journal/Jam _05/article01. html*, 2005.
[11] M. Mendoza and M. Rodríguez, "Aprendizaje centrado en el estudiante desde la planificación en investigación. CIENCIAMATRÍA, 6 (10), 560-572." 2019.
[12] J. J. C. Cantu, E. D. C. Ramirez, M. E. L. Almeida, E. S. I. Mendo, and J. M. Vergaray, "El aprendizaje autónomo en educación superior. Revisión sistemática," *Salud, Cienc. y Tecnol.*, no. 3, p. 391, 2023.
[13] F. Maffei, C. Neil, and N. Battaglia, "Herramientas para determinar estilos de aprendizaje basadas en Inteligencia artificial," in *XVII Congreso de Tecnología en Educación & Educación en Tecnología-TE&ET*, 2022.
[14] C. N. Suárez Guerrero, "La Zona de Desarrollo Próximo: categoría pedagógica para el análisis de la interacción en contextos de virtualidad," *Pixel-Bit*, 2004.
[15] S. L. Olivares Olivares and Y. Heredia Escorza, "Desarrollo del pensamiento crítico en ambientes de aprendizaje basado en problemas en estudiantes de educación superior," *Rev. Mex. Investig. Educ.*, vol. 17, no. 54, pp. 759–778, 2012.
[16] J. H. P. Prieto, "Secuencias didácticas: Aprendizaje y evaluación de competencias en educación superior," *Bordon*, vol. 63, no. 1. pp. 77–92, 2011.
[17] M. Neil, C. G., Battaglia, N., & De Vincenzi, "Diseño de competencia conducido por rúbricas," in *XXIX Congreso Argentino de Ciencias de la Computación*, 2023.
[18] M. P. García Sanz, "La evaluación de competencias en Educación Superior mediante rúbricas: un caso práctico," *Rev. Electrónica Interuniv. Form. del Profr.*, vol. 17, no. 1, 2014.
[19] E. Panadero, J. A. Tapia, and J. A. Huertas, "Rubrics and self-assessment scripts effects on self-regulation, learning and self-efficacy in secondary education," *Learn. Individ. Differ.*, vol. 22, no. 6, pp. 806–813, 2012.
[20] N. Battaglia, C. G. Neil, and M. De Vincenzi, "Software Engineering Competence-Based Learning in Collaborative Virtual Environments," in *EDUNINE 2021 - 5th IEEE World Engineering Education Conference: The Future of Engineering Education: Current Challenges and Opportunities, Proceedings*, 2021.
[21] J. B. Biggs, "Calidad del aprendizaje universitario," *Educ. Siglo XXI*, vol. 22, p. 272, 2004.
[22] O. D. L. Granizo, C. Neil, and C. S. C. Jiménez, "La inteligencia artificial en la educación y sus implicaciones: un mapeo sistemático de la literatura," *CONECTIVIDAD*, vol. 5, no. 1, pp. 49–66, 2024.
[23] C. Maffei, F., Neil, "Teorías de aprendizaje en el contexto de la Inteligencia Artificial," in *Congreso Nacional de Ingeniería Informática / Sistemas de Información 12º CoNaIISI. Universidad Tecnológica de Catamarca.*, 2024.
[24] C. Neil, N. Battaglia, and M. De Vincenzi, "Rubric-Driven Competency Development: A Case Study BT - Computer Science – CACIC 2023," P. Pesado, W. Panessi, and J. M. Fernández, Eds., Cham: Springer Nature Switzerland, 2024, pp. 350–365.
[25] C. G. Neil, *Marco conceptual para la definición, desarrollo y evaluación de competencias*, 1a ed. Ciudad Autónoma de Buenos Aires: Fundación Iberoamericana de Estudios Superiores, 2024.
[26] C. G. Neil, "Taller de definición, desarrollo y evaluación de competencias." https://uai.edu.ar/competencias
[27] C. G. Neil, N. Battaglia, and M. De Vincenzi, "La matriz de competencias como herramienta para orientar la escritura de resultados de aprendizaje," in *XVIII Congreso Nacional de Tecnología en Educación y Educación en Tecnología-TE&ET*, 2023.
[28] C. Neil, N. Battaglia, and M. E. D. V. Zemborain, "Marco metodológico para el diseño de rúbricas analíticas," *Edutec, Rev. Electrónica Tecnol. Educ.*, no. 80, 2022.

Estrategias pedagógicas y recursos didácticos para la enseñanza de las Ciencias de la Computación

Gustavo ASTUDILLO[1], Silvia BAST[1], Paula DIESER[1], Leandro CASTRO[1], Martín LOBOS[1], Yamila MINETTI[2] , Lucas DÍAZ[1]
[1] Departamento de Matemática/FCEyN/UNLPam, [2]FCH/UNLPam
astudillo@exactas.unlpam.edu.ar, silviabast@exactas.unlpam.edu.ar

RESUMEN

El proyecto de investigación aborda la creciente integración de la tecnología digital en la vida diaria y la importancia de enseñar Ciencias de la Computación en este contexto. Su objetivo principal es desarrollar nuevas estrategias y recursos didácticos para abordar esta demanda y responder a la pregunta: ¿Cómo enseñar Ciencias de la Computación en el contexto actual? Así, se enfoca en tres áreas: definir el perfil de las/los estudiantes según sus estilos de aprendizaje, proponer y validar métodos de evaluación, y desarrollar estrategias para incorporar aplicaciones en la enseñanza de Ciencias de la Computación.

Palabras clave: Estilos de aprendizaje, Evaluación de los aprendizajes, Estrategias innovadoras en la enseñanza, Ciencias de la Computación.

CONTEXTO

El presente trabajo se enmarca en las actividades del Grupo de Investigación y Desarrollo en Innovación Educativa (GrIDIE) y representa la continuidad y evolución del proyecto "Incorporación de estrategias innovadoras en los procesos de enseñanza y de aprendizajes de informática". La investigación actual profundiza en tres ejes fundamentales: (*i*) la caracterización y actualización del perfil de los estudiantes mediante el análisis de sus estilos de aprendizaje (EA), (*ii*) el desarrollo y revisión de nuevas propuestas de evaluación acordes al contexto educativo actual y (*iii*) la implementación y validación de estrategias innovadoras para la enseñanza de Ciencias de la Computación (CC), acompañadas del diseño y desarrollo de aplicaciones educativas que respondan a las necesidades pedagógicas identificadas.

El proyecto cuenta con el respaldo y financiamiento de la Facultad de Ciencias Exactas y Naturales (FCEyN - UNLPam) y ha sido acreditado para su ejecución durante el período 2024-2027 (Res 16/24 CD FCEyN - UNLPam).

1. INTRODUCCIÓN

La creciente digitalización de la sociedad contemporánea ha creado nuevos desafíos y oportunidades en el ámbito educativo, particularmente en la enseñanza de las CC. Como señalan Llambi et al. (2023), estamos presenciando una doble transformación: la penetración de las tecnologías computacionales en prácticamente todos los aspectos de nuestra vida diaria, junto con la emergencia de conceptos como inteligencia artificial, algoritmos, programación, robótica y pensamiento computacional en el panorama educativo.

En este contexto, donde la tecnología y la informática se han vuelto pilares fundamentales del progreso social y económico, resulta imperativo desarrollar, implementar y evaluar metodologías innovadoras para la enseñanza y el aprendizaje de las CC.

Este desafío nos impulsa a explorar y generar propuestas y recursos didácticos que faciliten la comprensión de los conceptos fundamentales de las CC en el entorno educativo actual.

2. LÍNEAS DE INVESTIGACIÓN, DESARROLLO E INNOVACIÓN

La hipótesis central del proyecto plantea que es posible mejorar la enseñanza y el aprendizaje de las CC mediante estrategias innovadoras que integren contenidos, actividades y evaluación con recursos digitales. La incorporación de Inteligencia Artificial Generativa (IAG), Programación Tangible o Robótica Educativa (RE) podría aumentar la motivación y facilitar la apropiación de los conceptos. Así como, una evaluación diversificada y la adaptación de las estrategias didácticas a los distintos EA darían lugar a la personalización de los aprendizajes.

El proyecto se estructura en tres ejes: revisión de enfoques sobre EA, diseño de estrategias de evaluación e integración de TIC en la enseñanza, desarrollados en las siguientes subsecciones.

Estilos de aprendizaje

Los estilos de aprendizaje (EA) han cobrado relevancia desde la década de 1970, aunque su conceptualización y aplicación carecen de una definición unificada (Alonso et al., 1994; Villarreal-Fernández, 2023). Keefe (1988) los define como un conjunto de cualidades cognitivas, afectivas y fisiológicas que influyen en el aprendizaje, mientras que Honey & Mumford (1986) los describen como las actitudes y comportamientos que determinan la forma preferida de aprender.

Diversos instrumentos permiten identificar los EA (Cárdenas-Palomino et al., 2022). Entre ellos, se seleccionó y aplicó CHAEA (Alonso et al., 1994), pero los resultados llevaron a cuestionar la validez, planteando nuevas preguntas de investigación sobre su confiabilidad en el ámbito universitario hispanohablante y posibles adaptaciones. Para responderlas, se realizó una revisión sistemática y un análisis psicométrico con el objetivo de desarrollar una herramienta ajustada al contexto.

Evaluación

En el nivel universitario, la evaluación educativa es clave tanto para estudiantes como para docentes. Como proceso multifacético, la evaluación es una parte indisociable de la propuesta de enseñanza (Celman, 1998). Al planificarse y diseñarse desde el inicio, facilita el seguimiento del aprendizaje y la reflexión sobre métodos, recursos y estrategias utilizadas.

Existen distintos enfoques en la evaluación, pero en su dimensión pedagógica cumple una función formativa, ya que proporciona información valiosa para ajustar la enseñanza. De este modo, la evaluación se integra en el proceso formativo con el propósito de favorecer el aprendizaje y los logros del estudiantado (Anijovich, 2010).

El proyecto busca proponer enfoques innovadores que, integrados a la enseñanza, favorezcan el aprendizaje.

Las TIC en la enseñanza de las CC

Esta línea de trabajo se centra en la adaptación e implementación de estrategias de enseñanza en CC, abordando tres ejes: programación tangible, robótica educativa e integración de IAG en el aula. Además, se busca desarrollar herramientas de software que complementen la propuesta pedagógica, con especial énfasis en programación tangible y robótica educativa.

La Programación Tangible (PT) permite crear programas mediante la manipulación de objetos físicos que representan comandos del lenguaje, sin necesidad de un computador (Resnick et al., 2000). Se basa en la manipulación de objetos físicos que, cuando están conectados correctamente, generan una salida en entornos digitales o robots (Castro, 2023).

Por otro lado, la robótica educativa, con un enfoque centrado en el estudiante, fomenta la creación de objetos tangibles con significado personal (Vaillant, 2013). Esta requiere un Entorno de Desarrollo Integrado (EDI) para programar y ejecutar instrucciones en los dispositivos. Existen diversos EDIs basados en bloques, como Mixly, Visualino o ArduinoBlocks (Lobos et al., 2021), pero deben adaptarse a la propuesta didáctica. Dado que el grupo desarrolló una secuencia didáctica para la enseñanza de programación utilizando

RE, fue necesario diseñar un IDE específico para su mejor implementación.

En un proyecto previo, se diseñó y aplicó una metodología de evaluación de IDEs para RE, identificando requisitos clave para su adaptación a la propuesta didáctica (Lobos et al., 2022) y el posterior desarrollo de la aplicación.

En el ámbito de la programación tangible se está avanzando en ampliar la revisión sistemática sobre la temática (Castro, 2023; Castro et al., 2021) con el fin de desarrollar una propuesta enfocada en la enseñanza de nociones básicas de programación.

La Inteligencia Artificial Generativa (IAG) se ha integrado en este proyecto mediante chatbots avanzados (ChatGPT, Gemini, Claude, Copilot, DeepSeek) y herramientas especializadas para investigación (Elicit, Consensus, ResearchRabbit), así como plataformas para el diseño de recursos audiovisuales (Suno, Hedra, Gamma, Canva, Udio). Su uso plantea interrogantes sobre su integración en el aula (UNESCO, 2019), la gestión de sesgos (Ferrante, 2021; Zou & Schiebinger, 2018) y alucinaciones (Cárdenas Herrera, 2023).

3. RESULTADOS OBTENIDOS/ESPERADOS

Desde el proyecto se han generado producciones propias que, junto con resultados previos, consolidan nuevas estrategias y enfoques. A continuación, se presentan los resultados alcanzados organizados por línea de trabajo:

Se realizó una revisión sistemática sobre el análisis psicométrico del CHAEA para evaluar su validez y confiabilidad y se consideraron adaptaciones a contextos específicos. Los hallazgos evidenciaron la necesidad de ajustes en función del contexto y avances tecnológicos, lo que motivó un nuevo análisis psicométrico.

Para validar su contenido, se exploró el uso de IAG como jueces expertos y se trabajó en la adaptación lingüística del cuestionario con un grupo reducido de destinatarios.

Posteriormente, se realizó un análisis métrico de los ítems a partir de una muestra estudiantes universitarios que derivó en una versión reducida del cuestionario y organizada en cuatro EA. Su estructura factorial se analizó mediante técnicas de teoría clásica de los tests. Finalmente, se evaluó la fiabilidad de esta versión, obteniendo resultados comparables a los de otras adaptaciones del CHAEA.

En el marco de la cursada 2024 de Introducción a la Computación, se propusieron estrategias innovadoras de evaluación para favorecer la apropiación de conceptos y el desarrollo de habilidades. Inicialmente, se revisó la propuesta evaluativa con el equipo docente, lo que llevó a dos decisiones: socializar listas de cotejo junto a los ejercicios previos al segundo parcial para familiarizar al estudiantado con los criterios de evaluación y realizar devoluciones orales y prácticas tras los parciales. Además, se ajustaron los criterios de evaluación de las Guías Prácticas y se definieron tres niveles cualitativos de calificación en la lista de cotejo general. Al finalizar, se analizaron los resultados de las evaluaciones y la encuesta de opinión sobre la cursada, permitiendo reflexionar sobre el impacto de las estrategias implementadas y su posible ajuste en futuras ediciones.

En la línea de Programación Tangible (PT) se realizó una revisión sistemática artículos publicados entre 2022 y 2024, identificando avances y matices respecto a la definición inicial. Se confirmaron aspectos fundamentales de la PT, como su énfasis en la interacción en el mundo real, el uso de conocimientos cotidianos y su potencial para fomentar la exploración y experimentación.

Además, se identificaron aplicaciones específicas, como la adaptación de sistemas para personas con discapacidad visual, el uso de robots para enseñanza STEM y el desarrollo de entornos de programación con realidad aumentada.

Se está desarrollando un EDI para robótica educativa denominado *GrIDIE Blocks*. Esta aplicación surge como respuesta a las

limitaciones identificadas en una evaluación de entornos de desarrollo integrados para robótica educativa, realizada en el marco del estudio "La medición y evaluación de EDI para robótica educativa". Este análisis reveló deficiencias en cuanto a accesibilidad y adaptación a los principios didácticos necesarios para una enseñanza progresiva de la programación. Ante la necesidad de mejorar la experiencia educativa en el Taller de Introducción a la Programación (TIP), se inició el desarrollo de *GrIDIE Blocks*, un EDI diseñado para simplificar la enseñanza de la programación mediante un enfoque basado en la didáctica por indagación, facilitando el aprendizaje con bloques en lenguaje natural y desafíos progresivos. Se han logrado hitos clave como la adopción de la metodología Scrum, el establecimiento de una base sólida en la interfaz del EDI y la estructura de desafíos y la implementación de funcionalidades esenciales como el bloque "ejecutar", bloques en lenguaje natural y procedimientos siempre presentes. También se ha priorizado de manera colaborativa las historias de usuario para asegurar el cumplimiento de los objetivos educativos.

Para hacer uso efectivo de la IA en el contexto educativo, es fundamental desarrollar habilidades como el pensamiento crítico para la detección de sesgos y el análisis de soluciones, competencias digitales, trabajo en equipo y aprendizaje continuo, todas ellas clave en la enseñanza de CC. En esta línea, se testeó un conjunto de herramientas de IA generativa para su aplicación en el aula y en investigación, Este trabajo dio lugar a la organización y desarrollo de un Taller de Posgrado, y a actividades de extensión para la difusión y formación en IA.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de investigación está conformado por investigadores en diferentes etapas de su formación académica: dos de ellos desarrollando sus tesis doctorales en distintas fases de avance, una investigadora que completó su maestría e inició su formación doctoral, y tres integrantes más que continúan avanzando en sus estudios de posgrado.

Como parte del compromiso con la formación de nuevos investigadores, se han incorporado dos estudiantes de grado como asistentes de investigación, y se mantiene una política activa de incorporación de estudiantes interesados en iniciarse en tareas de investigación, contribuyendo así a la renovación y fortalecimiento continuo del equipo.

BIBLIOGRAFÍA

Alonso, C. M., Gallego, D. J., & Honey, P. (1994). *Los estilos de aprendizaje: Procedimientos de diagnóstico y mejora*. Mensajero Bilbao, España.

Anijovich, R. (2010). *La evaluación significativa*. Ediciones Paidós.

Cárdenas Herrera, J. (2023). Inteligencia artificial, investigación y revisión por pares: Escenarios futuros y estrategias de acción. *RES. Revista Española de Sociología*, *32*(4), 199.

Cárdenas-Palomino, F., Beltran, J. A., Cámaco, D. D. la C., & Rojas-Zuñiga, L. (2022). Identificación de los Estilos de aprendizaje de los Estudiantes Universitarios: Revisión de Literatura. *Sinergias Educativas*. https://doi.org/10.37954/se.vi.255

Castro, L. J. (2023). *Programación tangible en el ámbito educativo* [Tesis, Universidad Nacional de La Plata]. http://sedici.unlp.edu.ar/handle/10915/154655

Castro, L. J., Artola, V., Astudillo, G. J., & Bast, S. G. (2021). Hacia un estado del arte sobre programación tangible. *Actas JADICC*, 45-53.

Celman, S. (1998). ¿Es posible mejorar la evaluación y transformarla en herramienta de conocimiento? En *La evaluación de los aprendizajes en el debate didáctico contemporáneo* (Vol. 35, pp. 35-66). Paidós.

Ferrante, E. (2021). Inteligencia artificial y sesgos algorítmicos ¿Por qué deberían importarnos? *Nueva Sociedad, 294*, Article 294.

Honey, P., & Mumford, A. (1986). *Using your learning styles*. Peter Honey.

Keefe, J. W. (1988). *Profiling and Utilizing Learning Style. NASSP Learning Style Series*. Publication Sales, National Association of Secondary School Principals, 1904 Association Drive, Reston, VA (Stock No.

Lobos, M., Bast, S. G., & Astudillo, G. J. (2021). *Selección de un entorno de desarrollo integrado de robótica educativa para la aplicación de una secuencia didáctica propuesta para el aprendizaje de conceptos de programación*. XXIII Workshop de Investigadores en Ciencias de la Computación (WICC 2021, Chilecito, La Rioja). http://sedici.unlp.edu.ar/handle/10915/120934

Lobos, M., Bast, S. G., & Astudillo, G. J. (2022). Evaluación de calidad de Entornos de Desarrollo Integrado de Robótica Educativa. En E. A. Irrazábal, G. N. Dapozo, & P. Martínez López (Eds.), *Libro de actas: II Jornadas argentinas de didáctica de las Ciencias de la Computación 2022* (pp. 156-167). Fundación Sadosky-Universidad Nacional del Nordeste. http://repositorio.unne.edu.ar/xmlui/handle/123456789/50765

Resnick, M., Berg, R., & Eisenberg, M. (2000). Beyond Black Boxes: Bringing Transparency and Aesthetics Back to Scientific Investigation. *Journal of the Learning Sciences*, *9*(1), 7-30. https://doi.org/10.1207/s15327809jls0901_3

UNESCO. (2019). Beijing Consensus on Artificial Intelligence and Education. *Planning education in the AI era: Lead the leap*, 70. https://unesdoc.unesco.org/ark:/48223/pf0000368303

Vaillant, D. (2013). *Integración de TIC en los sistemas de formación docente inicial y continua para la Educación Básica en América Latina*. UNICEF Argentina. https://www.unicef.org/argentina/spanish/educacion_Integracion_TIC_sistemas_formacion_docente.pdf

Villarreal-Fernández, J. E. (2023). Cuestionario Honey-Alonso de estilos de aprendizaje (CHAEA). Propiedades psicométricas en estudiantes universitarios colombianos. *Psicogente*, *26*(50), 46-69. https://doi.org/10.17081/psico.26.50.6231

Zou, J., & Schiebinger, L. (2018). AI can be sexist and racist—It's time to make it fair. *Nature*, *559*(7714), 324-326. https://doi.org/10.1038/d41586-018-05707-8

IMPLEMENTACIÓN DE MQTT Y APP INVENTOR PARA EL CONTROL REMOTO DE ROBOTS EN ENTORNOS DE APRENDIZAJE INTERACTIVO

Néstor Adrián Balich, Franco Adrián Balich, Magali Bolivar,
María Florencia Rossi, Matías Montiel, Romina Galizio
CAETI - Centro de Altos Estudios en Tecnología Informática
Universidad Abierta Interamericana. Informática (UAI)
Montes de Oca 745. Ciudad Autónoma de Buenos Aires, Argentina.

{nestor.balich, francoadrian.balich}@uai.edu.ar
{MagaliFlorencia.BolivarCruz, MariaFlorencia.Rossi, Carlaromina.galizio,
MatiasNicolas.MontielTorres }@alumnos.uai.edu.ar

Resumen

En la actualidad, el acceso a la educación tecnológica se ha expandido significativamente gracias a las plataformas digitales y las herramientas de aprendizaje interactivo. Sin embargo, persisten desafíos en la enseñanza de la robótica y la programación, especialmente en entornos donde los estudiantes no tienen acceso directo a hardware especializado. Para abordar esta problemática, en este proyecto se desarrolla un sistema de control remoto de robots educativos basado en la implementación del protocolo MQTT y la plataforma App Inventor, permitiendo que los estudiantes interactúen con los robots desde cualquier parte del mundo. El sistema se fundamenta en el uso de una placa Wemos D1 ESP8266 equipada con wifi, adicionando un par de motores de corriente continua, ruedas para tracción diferencial y un sensor de ultrasonido, permitiendo el sensado del entorno, la movilidad autónoma y la teleopreración. La comunicación se gestiona mediante un servidor MQTT, el cual facilita la interacción en tiempo real entre el robot y la aplicación móvil desarrollada en App Inventor. Además, la integración de una cámara web en la estación de pruebas posibilita la observación remota del entorno, proporcionando una experiencia más inmersiva para los estudiantes. El objetivo de este proyecto es democratizar el acceso a la educación en robótica, brindando a jóvenes estudiantes la posibilidad de realizar experimentos y pruebas con robots de manera remota. Al permitir el control interactivo desde cualquier ubicación con conexión a Internet, la utilización de herramientas libres, el desarrollo de un robot impreso en 3D de bajo costo que facilita su fabricación. este enfoque abre nuevas oportunidades para la enseñanza de la programación, la automatización y la inteligencia artificial en el ámbito educativo.

Palabras clave: Móvil, Robot, Wemos, Dispositivos Móviles, IoT, MQTT, App Inventor.

Contexto

El presente proyecto se desarrolla en el marco de las líneas de investigación del Laboratorio de Robótica Física e Inteligencia Artificial (LRFIA) de la Universidad Abierta Interamericana (UAI), en el Centro de Altos Estudios en Tecnología Informática (CAETI) con un enfoque en la aplicación de tecnologías emergentes en robótica y automatización educativa. En este proyecto participan docentes y alumnos tanto de sede Centro y Castelar (ambas en la provincia de Buenos Aires). En un mundo donde el acceso a equipos físicos puede ser una limitación para muchos estudiantes, este trabajo busca ampliar las oportunidades de aprendizaje mediante el uso de herramientas digitales que permitan la interacción remota con robots en entornos educativos.

1. Introducción

El control remoto de robots ha avanzado con el Internet de las Cosas (IoT) y el desarrollo de interfaces accesibles. La elección de protocolos de comunicación eficientes y plataformas de desarrollo intuitivas es clave para garantizar una interacción fluida.

El protocolo MQTT (Message Queuing Telemetry Transport) se destaca por su bajo consumo de ancho de banda y arquitectura ligera, permitiendo una comunicación eficiente en redes con conectividad limitada. Su modelo de publicación/suscripción optimiza la interacción entre el robot y la aplicación, asegurando respuesta en tiempo real [1][2].

Para la interfaz de usuario, se utilizó MIT App Inventor, una plataforma de desarrollo visual que permite crear aplicaciones móviles sin conocimientos avanzados de programación. Su entorno basado en bloques facilita el diseño de interfaces intuitivas, ideal para proyectos educativos [3].

El robot utiliza una placa Wemos como controlador de motores mediante un driver H y sensores, gestionar la comunicación recibiendo comandos a través de MQTT y la conexión a wifi, como así también el algoritmo de control del robot. Wemos es compatible con el laboratorio IDE de Arduino siendo una opción flexible y eficiente para automatización y robótica, con compatibilidad con sensores de ultrasonido para mejorar precisión y autonomía [4].

Este proyecto integra un servidor MQTT con una aplicación en App Inventor para controlar un robot impreso en 3D. Esta solución de bajo costo y escalable permite que estudiantes de cualquier parte del mundo interactúen con robots en entornos de aprendizaje remoto.

2. Líneas de Investigación y Desarrollo

Las principales líneas de investigación y desarrollo de este trabajo son:

1. **Implementación de un servidor MQTT:** Se configura un servidor MQTT utilizando Eclipse Mosquitto, garantizando una comunicación eficiente y segura entre la aplicación móvil y el robot. Se analizan aspectos clave como la escalabilidad, la gestión de clientes y la protección de datos para optimizar el rendimiento del sistema en entornos de acceso remoto.

2. **Desarrollo de una interfaz de control remoto:** Se diseña una aplicación móvil utilizando MIT App Inventor, una plataforma orientada a la democratización del desarrollo de software. Esta herramienta permite que los usuarios, sin necesidad de conocimientos avanzados en programación, puedan crear aplicaciones interactivas para el control del robot de forma intuitiva y accesible.

3. **Diseño y fabricación del robot:** El chasis del robot es desarrollado mediante impresión 3D, permitiendo una fabricación modular y de bajo costo. Además, se integran componentes electrónicos esenciales, como motores, sensores y una placa microcontroladora Wemos y Arduino, optimizando el diseño para facilitar la experimentación y la adaptación del hardware a diferentes necesidades educativas. [5]

4. **Integración de sistemas:** Se establece la conexión entre el servidor MQTT, la aplicación móvil y el robot, logrando un sistema de control remoto eficiente y funcional. Este proceso implica la configuración de la comunicación bidireccional, permitiendo que los usuarios envíen comandos al robot y reciban

retroalimentación en tiempo real a través de la interfaz móvil.

3. Resultados Obtenidos/Esperados

3.1 Resultados Obtenidos

La implementación del sistema de control remoto del robot ha demostrado ser funcional y eficiente, integrando con éxito un servidor MQTT, una aplicación móvil en MIT App Inventor y un robot basado en Arduino.

Comunicación con MQTT:

Se configuró un servidor MQTT utilizando Eclipse Mosquitto, estableciendo una comunicación confiable entre la aplicación móvil y el robot. Para ello, se utilizó la extensión UrsAI2MQTT en App Inventor, lo que permitió gestionar la publicación y recepción de mensajes de control. La implementación de MQTT ha demostrado ser una solución eficiente en aplicaciones de robótica, debido a su bajo consumo de ancho de banda y capacidad de comunicación en tiempo real. [1][6]

El proceso de integración de MQTT consistió en los siguientes pasos:

1. **Configuración de la Conexión:**
   - Se utilizó el bloque Conectar con la opción CleanSession activada, asegurando conexiones limpias y evitando mensajes retenidos.
   - Se verificó el estado de la conexión antes de proceder con el envío de comandos al robot.

2. **Publicación de Mensajes:**
   - Se definió un Topic denominado control/robot, donde la aplicación envió instrucciones de movimiento y comandos al robot.
   - Se implementó una verificación visual para confirmar el envío exitoso de los mensajes.

3. **Recepción de Mensajes:**
   - Se empleó la función PublishedReceived para recibir comandos en el tema control/robot.
   - La aplicación interpretó los mensajes y mostró la confirmación en pantalla, garantizando una comunicación bidireccional fluida. [7]

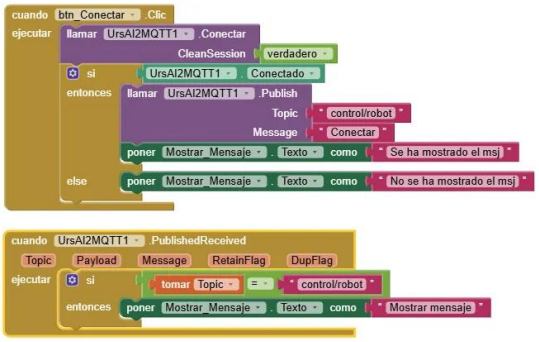


Figura 1. Ejemplo del boton de conexion de MQTT mediante la extension a UrsAI2MQTT en App Inventor

Desarrollo de la Interfaz de Control

Se diseñó una interfaz de usuario en App Inventor, incorporando botones táctiles para el control de movimiento y una sección de visualización de datos de sensores en tiempo real. La plataforma App Inventor, gracias a su programación basada en bloques, facilitó la implementación de la interfaz, permitiendo un desarrollo rápido e intuitivo sin necesidad de conocimientos avanzados en programación [3].



Figura 2. Visualización de la interfaz de la aplicación de App Inventor.

Diseño y Fabricación del Robot

El chasis del robot fue desarrollado mediante impresión 3D, lo que permitió un diseño modular y adaptable a futuras mejoras. Se integraron componentes electrónicos clave, incluyendo motores, sensores ultrasónicos y una placa Arduino, garantizando una estructura funcional y eficiente.

Desempeño del Sistema

- **Tiempo de respuesta promedio:** < 100 ms, permitiendo un control en tiempo real del robot.
- **Fiabilidad del sistema:** Comunicación estable a través de MQTT, con un 98% de éxito en la transmisión de mensajes.

3.2 Resultados Esperados

El siguiente paso en la evolución del proyecto es su expansión y optimización para mejorar la experiencia de aprendizaje remoto y la interacción con los robots. Para ello, se plantean los siguientes desarrollos:

- **Escalabilidad del sistema:** Se proyecta la implementación del sistema en al menos 10 robots adicionales, permitiendo que múltiples estudiantes puedan realizar pruebas y experimentos simultáneamente, fomentando un entorno de aprendizaje más dinámico y accesible.

- **Integración de cámaras para monitoreo remoto:** Se incorporarán cámaras web en la estación de pruebas, proporcionando a los usuarios una vista en tiempo real del robot y su entorno. Esto permitirá una mejor supervisión y control de los experimentos, facilitando una experiencia más inmersiva para los estudiantes.

Desarrollo de una mesa de pruebas interactiva: Se diseñará e implementará una mesa de pruebas equipada con sensores y marcadores, optimizada para la interacción con los robots. Esta infraestructura permitirá realizar experimentos en diferentes escenarios, mejorando la precisión y calidad del aprendizaje remoto.

Con estas mejoras, el proyecto busca consolidarse como una solución escalable y accesible para la educación en robótica, permitiendo a estudiantes de cualquier parte del mundo interactuar con robots en tiempo real y mejorar sus habilidades en programación y automatización. [8] [9] [10]

4. Formación de Recursos Humanos

- 2 docentes especialistas en IoT y robótica e impresión 3d, encargado de la supervisión técnica y metodológica del proyecto.
- 4 estudiantes de ingeniería en sistemas, responsables del desarrollo de la aplicación móvil, la integración de los sistemas electrónicos, la configuración y optimización del servidor MQTT y el desarrollo y optimización del chasis del robot mediante impresión 3D.

5. Bibliografia

[1] T. Mikolajczyk, H. Fuwen, L. Moldovan, A. Bustillo, M. Matuszewski, y K. Nowicki, “Selection of machining parameters with Android application made using MIT App Inventor bookmarks” Procedia Manufacturing, vol. 22, pp. 172–179, 2018. Disponible en: [Google Scholar] [CrossRef].

[2] T. Munasinghe, E. W. Patton, y O. Seneviratne, “IoT Application Development Using MIT App Inventor to Collect and Analyze Sensor Data,” en 2019 IEEE International Conference on Big Data, 2019. doi: 10.1109/BigData47090.2019.9006203.

[3] P. B. Moura Oliveira, “Teaching automation and control with App Inventor applications,” en 2015 IEEE Global Engineering Education Conference (EDUCON), Tallin, Estonia, 2015, pp. 879–884. doi: 10.1109/EDUCON.2015.7096077.

[4] G. A. Parra Bajaña, “Análisis comparativo de las herramientas App Inventor y Flutter para el desarrollo de aplicaciones móviles, caso desarrollo de aplicación de pedidos,” 2024. [En línea]. Disponible en: https://dspace.utb.edu.ec/handle/49000/15664

[5] J. Gómez y J. C. Macías, “Diseño e implementación de un robot paralelo 2RR para aplicaciones de impresión 3D,” 2022. [En línea]. Disponible en: http://hdl.handle.net/20.500.11912/10968

[6] E. González y J. Pérez, “Remote control of educational robots using MQTT and mobile applications,” Journal of Robotics and Automation, vol. 7, no. 3, pp. 45–58, 2021. doi: 10.1234/jra.2021.0345.

[7] A. Top y M. Gökbulut, “Android Application Design with MIT App Inventor for Bluetooth Based Mobile Robot Control,” 2021. doi: 10.21203/rs.3.rs-628321/v1.

[8] M. F. M. S., A. Selvakumar, y D. Schilberg, “Stereo Vision Based Robot for Remote Monitoring with VR Support,” 2024. doi: 10.35940/ijeat.A1016.1291S319.

[9] C. A. Valencia Hernández, “Metodología para un sistema de visión artificial apoyado con realidad aumentada en base a marcadores para un robot móvil,” 2015. [En línea]. Disponible en: https://repositorio.itm.edu.co/handle/20.500.12622/88.

[10] T. Mikolajczyk, H. Fuwen, L. Moldovan, A. Bustillo, M. Matuszewski, and K. Nowicki, “Selection of machining parameters with Android application made using MIT App Inventor bookmarks,” in Procedia Manufacturing, Elsevier B.V., 2018, pp. 172–179. doi: 10.1016/j.promfg.2018.03.027.

Incorporando el Análisis Sintáctico LR Canónico y el Análisis LALR a la herramienta JFlap

Alejandro Younes
alejandro.younes@est.fi.uncoma.edu.ar

Gerardo Parra
gparra@fi.uncoma.edu.ar

Grupo de Investigación en Lenguajes e Inteligencia Artificial
Departamento de Teoría de la Computación - Facultad de Informática
UNIVERSIDAD NACIONAL DEL COMAHUE

Resumen

El estudio de compiladores y de procesadores de lenguajes es esencial en la formación en informática. La tecnología de compiladores tiene un impacto profundo en numerosas áreas de las ciencias de la computación. En los cursos de compiladores, el análisis sintáctico es una de las fases más complejas dado que abarca un extenso contenido teórico-práctico. Existen algunas herramientas didácticas, como por ejemplo JFLAP, que facilitan la introducción de ciertos metódos de análisis sintáctico pero que, por otra parte, tienen ciertas limitaciones.

Esta propuesta se enmarca en una línea de investigación destinada a desarrollar recursos didácticos y evaluar su efectividad. El objetivo principal de esta línea es desarrollar recursos educativos y didácticos para colaborar en la enseñanza de conceptos avanzados de lenguajes de programación y de compiladores e intérpretes.

Palabras Clave: Educación en Ciencias de la Computación, Procesadores de Lenguajes, Análisis Sintáctico, Compiladores.

Contexto

El Grupo de Investigación en Lenguajes e Inteligencia Artificial (GILIA), de la Facultad de Informática de la Universidad Nacional del Comahue (UNCo) promueve el estudio de diversos temas de interés, entre los cuales se encuentran los procesadores de lenguajes. Esta línea de investigación y desarrollo se enmarca en el contexto de este grupo.

En particular, la línea de investigación se desarrolla en el ámbito del proyecto de investigación de la Facultad de Informática denominado *Modelos Formales, Agentes Inteligentes y Aplicaciones para la Enseñanza de las Ciencias de la Computación* (código 04/F022). El proyecto se inició en Enero de 2022 y tiene una finalización prevista para Diciembre de 2025.

1. Introducción

El diseño y construcción de compiladores proporciona la base necesaria para comprender desde otro punto de vista, las características de los diferentes lenguajes de programación. La tecnología de compiladores, en particular la parte de análisis, es necesaria para procesar la entrada de todo lenguaje, natural o formal, de cualquier sistema informático. Además, la relación entre el rendimiento de un sistema informático y la tecnología de compiladores es tan estrecha que los compiladores se utilizan usualmente como herramienta para evaluar conceptos de arquitecturas de computadoras antes de construir la arquitectura definitiva de

un nuevo procesador[1]. Por ello, los cursos de compiladores constituyen un punto importante en los programas de estudio de las carreras de Ciencias de la Computación. Estos cursos permiten a los estudiantes entender, entre otros conceptos, cómo el código fuente de sus programas va atravesando sucesivas etapas y transformaciones hasta arribar a la representación en código objeto o código destino.

Los cursos de compiladores tienden a ser algo complejos ya que se requiere que los estudiantes estén familiarizados con conceptos avanzados de los lenguajes de programación, teoría de la computación, diseño de algoritmos, estructuras de datos y arquitectura de computadoras, entre otros. Cada una de las fases de compilación representa una unidad didáctica independiente, con temas en los que abundan los conceptos abstractos. Los estudiantes usualmente reciben una comprensión básica de los compiladores a través de materiales teóricos, mientras que los ejercicios prácticos se dedican principalmente a algoritmos, estructuras de datos y construcciones teóricas[6]. Chesñevar y otros[4] sugieren, basándose en muchos años de experiencia docente, que la mayoría de los estudiantes inicialmente no se sienten motivados por el estudio de los lenguajes formales y teoría de la computación. Los autores creen que la razón de la falta de entusiasmo no es sólo la complejidad de los temas, sino también que los estudiantes suponen que estos temas están más relacionados con las matemáticas cuando, en realidad, corresponden a temática de ciencias de la computación. Por todo ello, es razonable suponer que la enseñanza de compiladores es también una tarea difícil para los profesores [17].

Los métodos de análisis sintáctico utilizados en compiladores pueden ser clasificados como descendentes o ascendentes. Tal como sugieren sus nombres, los métodos descendentes intentan construir, de manera implícita o explícita, árboles de derivación desde la raíz hacia las hojas, mientras que los métodos ascendentes comienzan desde las hojas y avanzan hasta la raíz.

Existen muchas técnicas de implementación específicas para el análisis sintáctico, como el método de análisis descendente *LL*(1) y los métodos de análisis ascendente LR Simple o *SLR*, *Lookahead* LR o *LALR* y LR canónico[1,2]. Básicamente, en todos ellos se requiere el cálculo de determinados conjuntos de símbolos, la construcción de tablas de análisis y el análisis de ciertas cadenas específicas. El cálculo de conjuntos como *Primeros*, *Siguientes* y la colección canónica de ítems LR(0)/LR(1), además de ser una carga de trabajo importante, al ser realizada usualmente de manera manual también es fácilmente propensa a tener errores. El análisis de cadenas, aunque sean de una longitud reducida, generalmente necesita una cantidad no despreciable de pasos de análisis. Por supuesto, un error en un paso provoca que el resto del análisis sea erróneo[18].

Existen varias herramientas didácticas con el fin de mejorar la enseñanza tradicional de análisis sintáctico y temas de compiladores, entre ellas, PAVT[15], EllErre[16], BURGRAM[5], ComVis[7,8], JSMachines[9] y JFLAP[11]. La herramienta JFLAP se utiliza en casi todo el mundo en cursos de teoría de la computación y de compiladores[12]. Para una gramática dada, se puede crear en JFLAP una tabla de análisis LL(1) de análisis descendente y una tabla de análisis SLR(1) de análisis ascendente. Además de construir de manera automática los conjuntos utilizados para crear dichas tablas de análisis: *Primeros*, *Siguientes* y en el caso del análisis SLR(1), la *colección canónica de ítems LR(0)*. Los estudiantes pueden realizar el análisis LL(1) y SLR(1) para cadenas dadas usando estas tablas y seguir el

proceso de análisis visualmente mediante la construcción paso a paso del árbol de derivación y/o la tabla de derivación. Varios estudios sobre el uso de JFLAP nos permiten notar una mejora significativa en la reducción de dificultad de los cursos de autómatas, lenguajes formales y compiladores[3, 13, 14].

El método de análisis ascendente SLR(1) no es lo suficientemente poderoso como para reconocer lenguajes generados por ciertas gramáticas que, sin ser ambiguas, pueden producir resultados conflictivos en la tabla de análisis sintáctico SLR(1). Para solucionar este problema, en algunos casos, es posible usar los métodos LALR(1) o LR(1). El método LR canónico o simplemente LR(1), hace uso de símbolos adicionales, denominados símbolos de *lookahead* y utiliza un extenso conjunto de elementos denominados ítems LR(1). El método lookahead-LR o LALR(1), se basa en los conjuntos de elementos LR(0) y posee mucho menos estados que los analizadores sintácticos basados en los ítems LR(1). Al introducir los símbolos de lookahead cuidadosamente en los ítems LR(0), podemos manejar muchas más gramáticas con el método LALR(1) que con el método SLR(1), y construir tablas de análisis sintáctico que no sean más grandes que las tablas SLR(1).

Una limitación importante de la herramienta JFLAP es su carencia de soporte para el análisis LALR(1) y el análisis LR canónico. Este último método de análisis es una característica muy deseable no sólo por su potencial uso en el curso de compiladores al ser el método de análisis ascendente más poderoso, sino también por sus implicaciones en la futura vida profesional de los estudiantes. Es importante destacar que, el análisis LR canónico constituye la técnica más general para construir una tabla de análisis sintáctico LR a partir de una gramática y por ello permite reconocer prácticamente todas las construcciones de los lenguajes de programación para las que se pueden diseñar gramáticas libres de contexto[2]. Otra limitación de la herramienta es la ausencia de alguna opción que permita observar cómo se construyen los conjuntos de símbolos y la colección canónica de ítems LR. A pesar de la naturaleza *paso a paso* de JFLAP, no permite ejecutar de tal forma los algoritmos que se enseñan en el curso de compiladores para el cálculo de los conjuntos necesarios. Esto hace que el uso de JFLAP esté limitado a la verificación de los resultados de dichos cálculos y que no se pueda utilizar para aprender "paso a paso" los algoritmos empleados.

Todo lo expuesto en los párrafos anteriores y la importancia del tema nos motiva a proponer el diseño y desarrollo de una extensión de JFLAP que permita utilizar los métodos de análisis LR Canónico y LALR(1). Además, la propuesta incluye la opción de ejecutar gráficamente paso a paso cada uno de los algoritmos utilizados, poniendo énfasis en que la interfaz de usuario sea intuitiva y amigable.

El resto del artículo está organizado de la siguiente manera. A continuación, describimos la línea de investigación y desarrollo. En la sección 3, presentamos los resultados esperados. Finalmente, en la sección 4, comentamos aspectos relacionados a la formación de recursos humanos en el marco de esta línea, así como del proyecto de investigación en su conjunto.

2. Línea de investigación y desarrollo

El objetivo general del proyecto de investigación 04/F022 es desarrollar construcciones teóricas, modelos y agentes inteligentes que fortalezcan los procesos de enseñanza de las Ciencias de la Computación. A partir de este objetivo general, se han realizado numerosos trabajos de campo para

abordar la enseñanza de conceptos de ciencias de la computación, como por ejemplo, temas relacionados a las redes de computadoras y nociones fundamentales de teoría de la computación. Los estudios desarrollados hasta ahora aportan resultados alentadores en relación a la efectividad de los recursos elaborados[10, 19, 20, 21].

El trabajo presentado en este artículo se enmarca en una nueva línea de investigación y desarrollo denominada *Recursos Educativos para Procesadores de Lenguajes*. El objetivo principal de esta reciente línea de investigación es desarrollar recursos educativos y didácticos para colaborar en la enseñanza de conceptos avanzados de lenguajes de programación y de compiladores e intérpretes.

3. Resultados esperados

El objetivo de este trabajo es el diseño e implementación de una extensión de JFLAP que incorpore al mismo los analizadores LR Canónico y LALR(1) para el apoyo a la enseñanza y aprendizaje de análisis sintáctico ascendente en cursos de compiladores.

Específicamente, se espera identificar y analizar las principales herramientas existentes utilizadas en la enseñanza de análisis sintáctico, diseñar módulos reusables para desarrollar la extensión e implementar un prototipo de la herramienta.

Se planea utilizar el producto resultante en los cursos de compiladores de nuestra carrera, como un apoyo para la enseñanza de los métodos de análisis sintáctico. Esto permitirá introducir de una manera más didáctica cada uno de los algoritmos utilizados. Asimismo, se espera que los estudiantes puedan utilizar la herramienta para analizar el funcionamiento de los diferentes métodos de análisis con entradas de su elección.

4. Formación de Recursos Humanos

Uno de los objetivos del desarrollo de esta línea de Investigación es contribuir a la formación de recursos humanos en el campo de Educación en Ciencias de la Computación. En este sentido, dos integrantes del proyecto de investigación están finalizando la Maestría en Ciencias de la Computación de la Facultad de Informática, UNCo. Además, otros dos integrantes están cursando la Maestría en Enseñanza en Escenarios Digitales, que desarrollan de manera conjunta las Universidades Nacionales de Cuyo, Comahue, Patagonia Austral, Patagonia San Juan Bosco, San Luis, Chilecito y La Pampa.

Los integrantes alumnos del proyecto también desarrollan sus tesis en temas relacionados.

Por otra parte, actualmente se están desarrollando al menos tres tesis de Licenciatura en Ciencias de la Computación en temas de interés del grupo de investigación.

5. Bibliografía

[1] A.V. Aho. Compilers: Principles, Techniques, & Tools. Pearson/Addison Wesley, 2007.

[2] A.V. Aho, R. Sethi y J.D. Ullman. Compilers: Principles, Techniques, and Tools. Addison-Wesley series in computer science. Addison-Wesley, 1986.

[3] Ryan Cavalcante, Thomas Finley y Susan H Rodger. «A visual and interactive automata theory course with JFLAP 4.0». En: Proceedings of the 35th SIGCSE technical symposium on Computer science education. 2004, págs. 140-144.

[4] Carlos I Chesñevar, Maria P González y Ana G Maguitman. «Didactic strategies for promoting significant learning in formal languages and automata theory». En: ACM SIGCSE Bulletin 36.3 (2004), págs. 7-11.

[5] César García-Osorio, Carlos Gómez-Palacios y Nicolás García-Pedrajas. «A tool for teaching LL and LR parsing algorithms». En: ACM SIGCSE Bulletin 40.3 (2008), págs. 317-317.
[6] William G Griswold. «Teaching software engineering in a compiler project course». En: Journal on educational resources in computing (JERIC) 2.4 (2002), 3-es.
[7] Nenad Jovanović et al. «ComVIS—Interactive simulation environment for compiler learning». En: Computer Applications in Engineering Education 30.1 (2022), págs. 275-291.
[8] Nenad Jovanović et al. «Teaching concepts related to finite automata using ComVis». En: Computer Applications in Engineering Education 29.5 (2021), págs. 994-1006.
[9] JSMachines. url: https://jsmachines.sourceforge.net/machines (visitado 19-02-2025).
[10] Dolz, Daniel, et al. Recursos Educativos Desenchufados para la Enseñanza de las Ciencias de la Computación en la Escuela Secundaria. XV Congreso Nacional de Tecnología en Educación y Educación en Tecnología (TE&ET 2020). Neuquén, 2020.
[11] Susan H Rodger y Thomas W Finley. JFLAP: an interactive formal languages and automata package. Jones & Bartlett Learning, 2006.
[12] Susan H Rodger, Jinghui Lim y Stephen Reading. «Increasing interaction and support in the formal languages and automata theory course». En: ACM SIGCSE Bulletin 39.3 (2007), págs. 58-62.
[13] Susan H Rodger et al. «Increasing engagement in automata theory with JFLAP». En: Proceedings of the 40th ACM technical symposium on Computer science education. 2009, págs. 403-407.
[14] Daniel Rodríguez y Salvador Sánchez-Alonso. «A preliminary evaluation of the impact of using a visual tool in a compilers course». En: 2nd Workshop on Methods and Cases (MCCE'08). 2008, págs. 47-52.
[15] Somya Sangal et al. «PAVT: a tool to visualize and teach parsing algorithms». En: Education and Information Technologies 23 (2018), págs. 2737-2764.
[16] L. M. Schnorr, M. C. Miletto y Solórzano A. L. V. EllErre: an LR automata generator. 2020. url: https://github.com/schnorr/ellerre (visitado 19-02-2025).
[17] Srećko Stamenković y Nenad Jovanović. «A Web-based Educational System for Teaching Compilers». En: IEEE Transactions on Learning Technologies (2023).
[18] Decheng Wang y Yan Chen. «Visual teaching design of compile principle based on JFLAP». En: 2021 2nd International Conference on Artificial Intelligence and Education (ICAIE). IEEE. 2021, págs. 366-370.
[19] Dolz, D., Parra, G., Rodríguez, J. Diseño Participativo para desarrollar Recursos Educativos para la Enseñanza de las Ciencias de la Computación en la Escuela Secundaria. Una Máquina de Turing en la Escuela. WICC 2022. U. Champagnat. Mendoza, Argentina.
[20] Dolz, D., Parra, G., Rodríguez, J., Parra S. Desarrollo de Recursos Educativos para la Enseñanza de las Ciencias de la Computación en la Escuela Secundaria. WICC 2023. U. N. del Noroeste de la pcia. de Buenos Aires.
[21] Rodríguez, J, et al. Una Máquina de Turing en la Escuela. CACIC 2022. U. N. de La Rioja. La Rioja, Argentina.

INTERVENCIÓN ESTRATÉGICA A TRAVÉS DE VIDEOJUEGOS

Hoferek Silvia Rosana[1,3], Aranda Marcos[1,2], Bravo Luis [1], Gianotti Martín Gastón [1], Gónzalez Ricardo Elian [1]

[1] Universidad Siglo 21. Decanato de Ciencias Aplicadas.
srhoferek@gmail.com - marcos.aranda@ues21.edu.ar
[2] Universidad Nacional de Catamarca. Facultad de Tecnología y Ciencias Aplicadas.
[3] Universidad de la Cuenca del Plata. Facultad de Ingeniería, Tecnología y Arquitectura.

,

RESUMEN

Los videojuegos son herramientas de comunicación que estimulan aprendizajes informales y ubicuos. Su éxito depende del cumplimiento del objetivo y del compromiso de los jugadores. Sin embargo, se percibe una falta de videojuegos con árboles de decisión y modelado de roles que fomenten la autodeterminación y autoconocimiento juvenil. Para abordar este desafío, se propone el marco MDA (Mecánicas, Dinámicas y Estética) en el diseño y evaluación de un videojuego de simulación que motive a los jóvenes a reflexionar sobre sus actitudes y valores en la toma de decisiones. Actualmente, la investigación se encuentra en la recolección bibliográfica sobre videojuegos serios y su impacto en habilidades cognitivas. Se aplicará un estudio experimental con grupos control y tratamiento, donde la variable independiente será jugar el videojuego y la dependiente, la capacidad de toma de decisiones. Se espera contribuir con un modelo de intervención estratégica basado en videojuegos.

PALABRAS CLAVES: Videojuegos serios, Toma de decisiones, Aprendizaje ubicuo.

CONTEXTO

Los videojuegos son herramientas de comunicación que generan un entorno interesante en los jóvenes, promoviendo su aprendizaje independiente. Por medio de las tecnologías de la información y la comunicación los juegos han evolucionado hacia experiencias cada vez más inmersivas, favoreciendo el desarrollo de habilidades cognitivas y en la toma de decisiones. Si bien existen videojuegos que pueden ayudar en la toma de decisiones y el desarrollo de destrezas, se identifica una carencia de juegos que integren árboles de decisión y modelado de roles enfocados en la autodeterminación y el autoconocimiento de los jóvenes.

El objetivo de esta investigación es diseñar y evaluar un modelo de intervención estratégica basado en un videojuego de simulación. Actualmente, este trabajo se encuentra en la fase de recolección bibliográfica, explorando la evolución de los videojuegos y su impacto en la formación de habilidades cognitivas. Con esta investigación se busca aportar un modelo innovador para el aprendizaje de los jóvenes.

1. INTRODUCCIÓN

Los videojuegos son herramientas muy importantes en la educación y el desarrollo cognitivo. Su capacidad para fomentar el aprendizaje a través de la interacción y la simulación generan un impacto en la toma de decisiones [1], [2]. Diferentes estudios realizados recientemente han demostrado que los videojuegos pueden actuar como ejercicio cognitivo, fortaleciendo habilidades como la memoria, la atención y la resolución de problemas [3], [4].

El éxito de un videojuego educativo se mide a través de su capacidad para captar la atención y promover la adquisición de conocimientos [5]. En este contexto, la teoría MDA es un enfoque clave en el diseño de videojuegos que buscan influir en el comportamiento y las habilidades cognitivas [6].

A pesar de los diferentes avances en el tiempo, se ha identificado que los videojuegos carecen de árboles de decisión [7]. Debido a esta problemática, este trabajo propone el desarrollo de un videojuego de simulación basado en la teoría de la decisión y la incertidumbre, utilizando elementos de gamificación que ayuden a la reflexión y el análisis crítico [8] ,[9].

Este trabajo espera que los resultados contribuyan a la creación de un modelo de intervención estratégica que potencie la capacidad de decisión en los jóvenes mediante el uso de videojuegos diseñados con fines educativos.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Los objetivos para el desarrollo de este trabajo son:

- Diseñar y desarrollar un videojuego educativo basado en el marco MDA que permita a los jóvenes mejorar su capacidad en la toma de decisiones mediante la simulación de escenarios interactivos, integrando árboles de decisión y modelado de roles para fomentar la autodeterminación y el autoconocimiento.
- Evaluar el impacto del videojuego en la mejora de habilidades cognitivas y toma de decisiones a través de un estudio experimental con grupos de control y tratamiento, midiendo variables como el nivel de reflexión, análisis crítico y resolución de problemas en distintos contextos de incertidumbre.
- Analizar la efectividad del modelo de intervención estratégica basado en videojuegos mediante la recopilación y comparación de datos cuantitativos y cualitativos, identificando patrones de aprendizaje en los participantes, con el fin de optimizar futuras implementaciones en entornos educativos y de formación.

3. RESULTADOS ESPERADOS

Este trabajo espera diseñar y validar un modelo de intervención estratégica basado en videojuegos, incentivando la autorreflexión en jóvenes sobre sus decisiones. Se implementará un árbol de decisiones adaptable con machine learning, permitiendo flexibilidad ante cambios y considerando restricciones del juego y esquemas de valores para evaluar el impacto moral. Además, se realizará un análisis estadístico y evolutivo de videojuegos serios enfocados en la toma de decisiones, comparando su efectividad con el modelo propuesto.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo está conformado por siete integrantes: tres docentes de la Licenciatura en Informática, pertenecientes al Área de Ciencias Aplicadas de la Universidad Siglo 21, y cuatro alumnos de la misma institución. La participación estudiantil busca fomentar su formación en investigación y motivarlos en el desarrollo de productos de software innovadores con impacto social.

La diversidad del equipo radica en la presencia de docentes de distintas cátedras y estudiantes de diferentes años y modalidades de cursado. Esta composición

impulsa una educación de calidad alineada con los objetivos de desarrollo sostenible, promoviendo la formación de agentes de cambio capaces de adaptarse a un entorno tecnológico en constante evolución.

En cuanto a la formación académica del equipo, dos docentes están cursando un Doctorado en Informática y otro en Ingeniería, con líneas de investigación en desarrollo tecnológico con microprocesadores de última generación e inteligencia artificial. Su trayectoria incluye participación en congresos y publicaciones científicas.

5. BIBLIOGRAFÍA

[1] Y. del S. Barreto-Cabrera, A. Suárez Perdomo, and J. L. Castilla-Vallejo, "Perfiles de uso problemático de los videojuegos y su influencia en el rendimiento académico y los procesos de toma de decisiones en alumnado universitario," *Pixel-Bit. Revista de Medios y Educación*, vol. 69, p. 287, 2024.

[2] M. I. Gómez-León, "Eficacia de los videojuegos en la regulación emocional de los estudiantes," *Tecnología, Ciencia y Educación*, vol. 29, pp. 31-58, 2024.

[3] D. Carvajal, "Los videojuegos favorecen el desarrollo del cerebro y son una herramienta de aprendizaje efectiva," *Universidad Politécnica de Madrid*, 2024.

[4] A. Owen, "Scientists say that playing video games is like a gym workout for your brain," *The Scottish Sun*, 2024.

[5] P. Burelli and L. Dixen, "Playing With Neuroscience: Past, Present and Future of Neuroimaging and Games," *arXiv preprint arXiv:2403.15413*, 2024.

[6] G. Barbero, M. A. Gómez-Maureira, and F. F. J. Hermans, "Computational Thinking through Design Patterns in Video Games," *arXiv preprint arXiv:2407.03860*, 2024.

[7] "La influencia del uso de videojuegos educativos en el desarrollo cognitivo infantil," *593 Digital Publisher CEIT*, 2024.

[8] H. Guerrero Salas, M. Mayorga Morato, and O. Suárez de Antonio, *Teoría de la Decisión Aplicada. 2° Edición. Análisis de las Decisiones Bajo Incertidumbre, Riesgos, Teoría de Juegos y Cadena de Markov*, ECOE Ediciones, Bogotá, 2023.

[9] S. Ballesteros, "Investigadores de la UNED buscan prevenir demencias a través de los videojuegos," *UNED*, 2024.

Metodologías y recursos para la enseñanza de la programación: Evaluación con rúbricas, evaluación de calidad y el rol de la IA

Gladys N. Dapozo, Cristina L. Greiner, Raquel H. Petris, Ana M. Company, María C. Espíndola, Silvana Armana

Departamento de Informática. Facultad de Ciencias Exactas y Naturales y Agrimensura
Universidad Nacional del Nordeste. 9 de Julio Nº 1449. Corrientes. Argentina.
{gndapozo, cgreiner, rpetris, acompany, mcespindola}@exa.unne.edu.ar;
sil_armana@yahoo.com.ar

RESUMEN

Este proyecto de investigación estudia métodos y herramientas para mejorar la enseñanza de la programación en la universidad y niveles preuniversitarios. Se describen experiencias con rúbricas en Moodle y la perspectiva estudiantil sobre el uso de IA generativa como apoyo al aprendizaje en la universidad. Además, se presentan resultados sobre la aplicación de un modelo de calidad para evaluar herramientas software utilizadas en la enseñanza de la programación en niveles preuniversitarios.

Palabras clave: Evaluación con rúbricas. Modelo de calidad de herramientas sofware. IA generativa en la enseñanza de la programación.

CONTEXTO

Las líneas de I/D corresponden al proyecto 21F016 “Modelos, metodologías y recursos para el desarrollo del pensamiento computacional”, acreditado por la SGCyT-UNNE, correspondiente al periodo 2022-2025.

1. INTRODUCCIÓN

En los últimos años, en Argentina, se ha observado en las carreras de informática un notable incremento de su matrícula. En parte se debe a la aceleración de la virtualización provocada por la pandemia, al incremento de la salida laboral, la masificación de la modalidad de trabajo remoto, mejores retribuciones salariales para los trabajadores del sector, entre otras. Respecto de este último punto, cabe destacar que los sueldos de la industria del software duplican la Canasta Básica y los salarios del sector privado en general, según informe del Observatorio Permanente de la Industria del Software y Servicios Informáticos de la República Argentina (OPSSI) [1].

Por otra parte, la sociedad en la actualidad está más informada sobre la diversidad de aplicaciones posibles de la informática y el potencial de estudiar una carrera informática y conseguir un trabajo interesante y con buenas condiciones parece ser dato de la realidad compartida. Sin embargo, el aumento del interés por estudiar informática pone en tensión a las instituciones en relación con su infraestructura, equipamiento y disponibilidad de recursos humanos para atender a un número creciente de estudiantes en los primeros años [2].

En contextos de masividad las estrategias y metodologías de enseñanza y aprendizaje cobran especial importancia para lograr los objetivos de aprendizaje, sin descuidar la calidad académica. En particular, referido a la evaluación se considera que las rúbricas ofrecen ventajas que pueden contribuir a sistematizar los resultados y poder obtener una retroalimentación de los alumnos que permita detectar tempranamente las dificultades comunes e implementar acciones correctivas o tutorías dirigidas a los problemas más frecuentes.

LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Las líneas de investigación de este proyecto están enfocadas en:

Estrategias educativas que incorporen métodos y herramientas innovadoras para la enseñanza de la programación en las carreras de Informática.

Evaluación de métodos y herramientas para la enseñanza de programación y robótica para docentes de los niveles educativos no universitarios.

2. RESULTADOS OBTENIDOS/ESPERADOS

En la línea de estrategias educativas para la enseñanza de programación en carreras de Informática, en [3] se presentan los resultados de una experiencia de implementación de una rúbrica en Moodle para evaluar el logro de los objetivos de aprendizaje en una asignatura de primer año de una carrera de Informática con el propósito de tener una retroalimentación automatizada del desempeño de los estudiantes. La implementación de esta herramienta presenta en un contexto de masividad, las siguientes ventajas:

Desde el punto de vista del proceso de enseñanza y aprendizaje, el análisis de los resultados de la rúbrica brinda un panorama completo para identificar las dificultades de los estudiantes y habilitar acciones correctivas en el proceso de enseñanza.

Además, el estudiante recibe una devolución detallada y personalizada de su examen, en la cual se señalan los errores cometidos y las recomendaciones dadas por los docentes. En contextos de masividad esta tarea se convierte en agobiadora y extensa, mientras que en esta modalidad el alumno recibe la devolución en el aula virtual y la procesa y asimila con mayor efectividad.

Desde el punto de vista de la coordinación de un equipo docente numeroso, la aplicación de una rúbrica digitalizada se convierte en un ordenador de los criterios de evaluación dado que se acuerdan los criterios, los niveles y la ponderación de cada uno de ellos, en función de los objetivos de aprendizaje de la asignatura, cuya aplicación da como resultado la calificación del examen, minimizando las subjetividades y logrando mayor equidad para todos los alumnos.

Además, desde lo operativo, agiliza el registro de las calificaciones dado que Moodle permite exportar las calificaciones en por ejemplo una planilla de cálculo, evitando una transcripción manual de las calificaciones.

A futuro, para consolidar esta metodología se prevé realizar una encuesta a los estudiantes para evaluar el grado de satisfacción de los mismos con la implementación de esta herramienta. Del mismo modo se recabará la opinión de los docentes sobre este mecanismo para evaluar el aprendizaje de los estudiantes en los temas específicos de programación.

Además de la masividad de estudiantes, la aparición de la inteligencia artificial (IA) generativa ha impactado en la educación en Ciencias de la Computación. Las herramientas de inteligencia artificial, como ChatGPT y GitHub Copilot, han atraído una atención significativa por parte de los educadores en informática. Tanto los investigadores como los profesionales han descubierto que estas herramientas pueden generar soluciones correctas para una variedad de tareas introductorias a la programación y explicar con precisión el contenido del código. Dadas sus capacidades actuales y los probables avances en los próximos años, ¿habrá que realizar adaptaciones para garantizar que los estudiantes sigan aprendiendo bien? [4].

Atentos a esta situación, al finalizar el dictado 2024 de la asignatura Algoritmos y Estructuras de Datos I, se realizó una encuesta con el objetivo de indagar cómo utilizaron las herramientas de IA en su aprendizaje y su opinión sobre su eficacia. Los resultados se presentaron en [5]. La Fig. 1 evidencia que un alto porcentaje (83%) de los estudiantes reconoce haber utilizado una herramienta de IA durante el cursado de la asignatura,

aunque han encontrado dificultades para obtener soluciones alineadas con las pautas de la asignatura. Algunas de las dificultades señaladas se muestran en la Fig. 2.

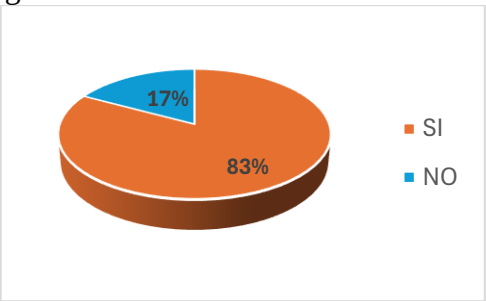


Figura 1: Uso de herramientas de IA durante el cursado de la asignatura

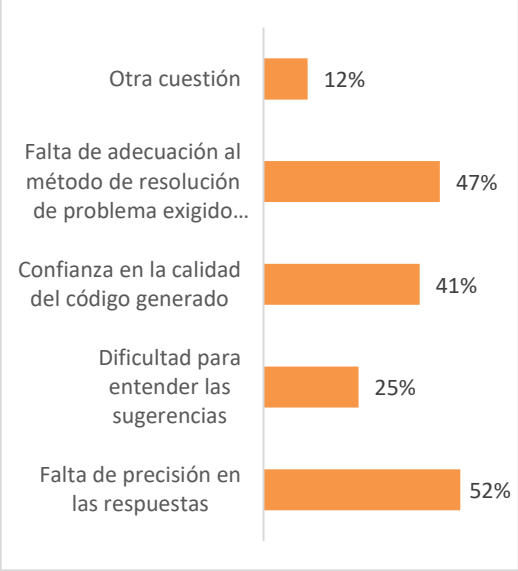


Figura 2: Principales desafíos de la IA para aprender programación

Consultados acerca de si considera que el uso de la inteligencia artificial será una habilidad esencial para futuros programadores, en la Fig, 3 se puede apreciar que la mayoría de los estudiantes (69%) considera que la IA será una habilidad esencial para futuros programadores.

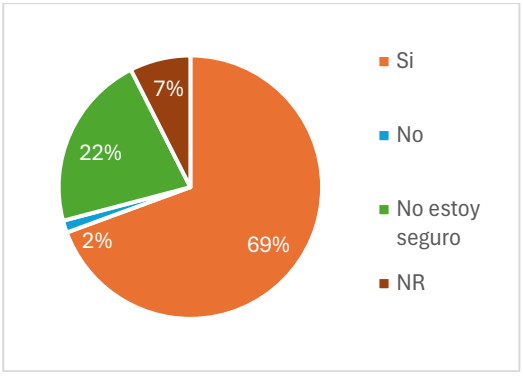


Figura 3: Importancia de la IA para los programadores

Esto subraya la necesidad de una instrucción formal que enseñe a formular preguntas adecuadas para alcanzar las soluciones esperadas. Como conclusión, se propone elaborar a futuro una metodología que integre las herramientas de IA en la enseñanza, buscando maximizar su uso sin comprometer el desarrollo del pensamiento computacional y la abstracción necesarios para una sólida formación en programación

En la línea de evaluación de métodos y herramientas para la enseñanza de programación para docentes de los niveles educativos no universitarios, se elaboró un marco de referencia para la evaluación de calidad de las herramientas utilizadas en la enseñanza de la programación [6]

Para la elaboración de la propuesta se tuvo en cuenta la norma ISO/IEC 25000, conocida como SQuaRE (Requisitos y Evaluación de Calidad de Productos de Software) una familia de normas que tiene por objetivo la creación de un marco de trabajo común para evaluar la calidad del producto software [7].

En primer lugar, para la definición del modelo de calidad, se consideraron las características y subcaracteristicas más apropiadas para esta evaluación, las que se muestran en la Figura 4.

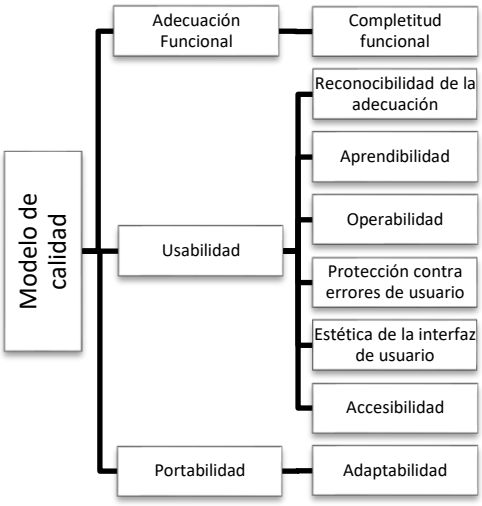


Figura 4. Modelo de calidad para la evaluación de herramientas de programación.

Luego se realizó la ponderación de cada una de las características en el modelo, según su relevancia para la evaluación de recursos educativos, según se muestra en la Tabla 1.

Tabla 1. Ponderación de las características

Característica	Ponderación
Adecuación Funcional	20%
Usabilidad	50%
Portabilidad	30%

A continuación, se definieron las métricas, basadas en la ISO/IEC 25023, para determinar el grado de cumplimiento de las características seleccionadas para el modelo de calidad propuesto [8], las que se muestran en la Tabla 2.

Tabla 2. Métricas para las características y subcaracteristicas del modelo de calidad

Subcaracterística	Métrica
C1. Adecuación funcional	
SB1.Completitud Funcional	M1. Completitud de la implementación
C2. Usabilidad	
SB1. Reconocibilidad de la adecuación	M1. Medida de idoneidad y reconocibilidad
SB2. Aprendibilidad	M1. Integridad de la guía de usuarios
	M2. Efectividad de la documentación del usuario o ayuda del sistema
SB3. Operabilidad	M1. Claridad del mensaje
SB4. Protección contra errores del usuario	M1. Prevención del uso incorrecto
SB5. Estética de la interfaz de usuario	M1. Personalización de la apariencia de la interfaz del usuario
	M2. Legibilidad adecuada
SB6. Accesibilidad	M1. Accesibilidad para usuarios con discapacidad
	M2. Idiomas admitidos
C3: Portabilidad	
SB1.Adaptabilidad	M1. Adaptabilidad en entorno software

Finalmente, se definió el proceso de evaluación del modelo propuesto, basado en la norma ISO/IEC 25040, que forma parte de la ISO/IEC 25000. Esta norma establece un proceso de evaluación estructurado en etapas, que se detallan en [9].

Se realizó la evaluación de calidad tomando como caso de estudio a tres herramientas de enseñanza de programación utilizados en los niveles educativos preuniversitarios: PilasBloques, Scratch y Gobstones. En la Tabla 3 se exponen los resultados de la evaluación de las características en cada uno de los productos software.

Tabla 3. Resultados de la evaluación de las Características

Caracte-rísticas	Ponde ración	PB	Scrat ch	Gobs-tones
Adecuación funcional	20%	18%	15%	12%
Usabilidad	50%	37%	36%	25%
Portabilidad	30%	30%	30%	30%
Total	100%	85%	82%	67%

Análisis de las características del modelo:

1-Adecuación Funcional: tomando como base la ponderación (20%), se observa en la Figura 5 el cumplimiento de esta característica. Pilas Bloques presenta un alto grado de cumplimiento de esta característica (18/20), lo que indica una sólida completitud funcional. Esto significa que cumple en gran medida con los requisitos necesarios para la enseñanza de la programación, proporcionando una amplia gama de funcionalidades. Scratch, con un nivel menor, sugiere un cumplimiento con algunas limitaciones de esta característica, pero en mejor situación que Gobstones, que con un 60% presenta el menor grado de cumplimiento de la adecuación funcional. En la Figura 5 se muestra la representación del valor para cada herramienta.

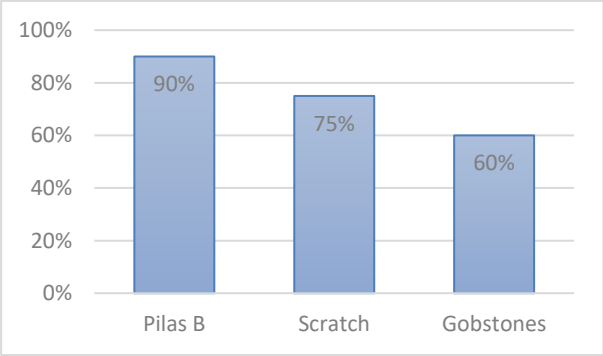


Figura 5. Adecuación Funcional

2-**Usabilidad**: tomando como base la ponderación 50% asignada al modelo, se observan gráficamente los resultados en la Figura 6.

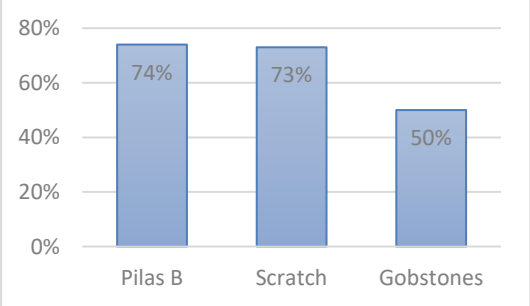


Figura 6. Usabilidad

Pilas Bloques y Scratch muestran valores similares de cumplimiento de esta característica lo que indica que ambas herramientas ofrecen una buena experiencia de usuario. Gobstones con un cumplimiento bastante menor, sugiere que enfrenta mayores desafíos en términos de facilidad de uso, lo que podría afectar la efectividad y la satisfacción del usuario.

3- **Portabilidad**: Tomando como base la ponderación 30% asignada al modelo, el resultado obtenido indica que Pilas Bloques, Scratch y Gobstones alcanzan un cumplimiento del 100%. Esto significa que las herramientas evaluadas son altamente portables y accesibles en una variedad de plataformas y dispositivos. Esta alta portabilidad asegura que los usuarios puedan trabajar en sus proyectos desde diferentes entornos sin preocuparse por problemas de compatibilidad, y que el aprendizaje realizado con estas herramientas se mantenga en el tiempo.

La validación del modelo de calidad en la evaluación de herramientas software de enseñanza de la programación ha demostrado su eficacia dado que ha permitido un análisis comparativo sistemático de las características de estas herramientas, en particular en aspectos relevantes para la enseñanza, como son la adecuación funcional, la usabilidad y la portabilidad. La información resultante de la evaluación permitirá guiar a los docentes en la selección de las herramientas más adecuadas para sus necesidades, reduciendo la incertidumbre en el proceso de elección mediante la aplicación de criterios claros y específicos.

El modelo es versátil y puede adecuarse para la evaluación de otras herramientas y otras características que resulten de interés, siguiendo el mismo esquema que se muestra en este trabajo.

A futuro, se propone evaluar estas herramientas de programación desde la perspectiva de los usuarios, tanto

docentes como estudiantes, siguiendo las pautas de la ISO/IEC 25022 sobre Calidad en Uso. Esto podrá enriquecer la información sobre estas herramientas dado que surgirá de experiencias concretas de docentes o estudiantes al usar estos entornos. También podrá asociarse a esta evaluación el nivel educativo, edad de los estudiantes, tipo de actividades, entre otras posibles.

FORMACIÓN DE RECURSOS HUMANOS

En este proyecto participan cinco docentes investigadoras, entre ellas dos tesistas de posgrado que finalizaron la Maestría en Tecnologías de la Información de la UNNE en el año 2024.

3. REFERENCIAS

[1] Los sueldos de la industria del software duplican la Canasta Básica y los salarios del sector privado en general, https://cessi.org.ar/wp-content/uploads/2023/06/Infografia-Salarios-Software-enero-2023.pdf , consultado por última vez el 20/03/2024

[2] Quartulli, D., Curcio, J.: ¿Por qué estudiamos informática? Indagación sobre trayectorias universitarias: instituciones, estudiantes, género y trabajo. Fundación Sadosky, Buenos Aires (2023)

[3] Gladys Dapozo, Cristina Greiner, Raquel Petris, Ana María Company, María Cecilia Espíndola, Silvana Armana, María Isabel Sánchez. Use of rubrics in Moodle to assess learning in programming in massive university contexts Proceedings of the 19th Latin American Conference on Learning Technologies (LACLO 2024) *(en prensa)*

[4] S. Lau, P. J. Gou, “From "Ban It TillWe Understand It" to "Resistance is Futile": How University Programming Instructors Plan to Adapt as More Students Use AI Code Generation and Explanation Tools such as ChatGPT and GitHub Copilot”, in Proceedings of the 2023 ACM Conference on International Computing Education Research, New York, NY, USA, 2023

[5] Gladys Dapozo, Cristina Greiner, Raquel Petris, Ana María Company, María Cecilia Espíndola, Uso de IA en programación inicial desde la perspectiva de los estudiantes. Jornadas Argentinas de Didáctica de las Ciencias de la Computación (JADiCC 2024), Facultad de Ciencias Exactas Físico-Químicas y Naturales. Universidad Nacional de Río Cuarto.

[6] María Cecilia Espíndola. Evaluación de calidad de herramientas utilizadas en la enseñanza de la programación basada en ISO 25000. Tesis de maestría. Disponible en: http://repositorio.unne.edu.ar/handle/123456789/30375

[7] ISO/IEC 25010. Modelo de calidad. https://iso25000.com/index.php/normas-iso-25000/iso-25010

[8] ISO, “ISO, ISO/IEC 25023:2016 (Ed 1, 2016). Systems and software engineering - Systems and software”, Available: ttps://www.iso.org/standard/35747.html

[9] ISO, “ISO25000 Calidad de software y datos”. Available: https://iso25000.com/index.php/normas-iso-25000/iso-25040

Plan estratégico para el fortalecimiento de las competencias blandas en el currículo del programa Ingeniería en Software, Informática, Ciencias Computacionales y afines

María V. Poliche[1], Claudia M. Herrera[1], Carolina I. Chayle[1], Manuel Alexander Valbuena Henao[2], Silvana Lorena Vallejo Córdoba [2], Juan David Tamayo Quintero[2]
[1] Departamento de Informática- Facultad de Tecnología y Ciencias Aplicadas - Universidad Nacional de Catamarca
[2] Departamento de Informática- Facultad de Tecnología y Ciencias Aplicadas - Universidad Nacional de Catamarca
vpoliche@tecno.unca.edu.ar cm_herrera@tecno.unca.edu.ar cchayle@tecno.unca.edu.ar

RESUMEN

El distanciamiento entre el sistema de Educación Superior y el mercado laboral se ha intensificado de manera significativa con el paso del tiempo. Además de considerarse como responsable de la desmejora en la productividad, la competitividad y la empleabilidad; esta falta de acoplamiento entre la academia y la industria ocasiona brechas de capital humano que se evidencian con la falta de trabajadores provistos con las habilidades necesarias para suplir las demandas del sector productivo. En este escenario, el marco de desajuste de habilidades existente entre la formación disponible y aquella requerida por el mercado laboral de la industria de las Tecnologías de la Información (TI), también se hace evidente en el ámbito global. Bajo esta perspectiva, la explicación para la brecha de habilidades existente en la oferta laboral disponible para el sector TI puede obedecer a la carencia de las habilidades integrales que conjugan aspectos técnicos propios de las ocupaciones del área disciplinar, denominadas competencias duras, y aspectos generales relacionados con habilidades sociales y de comportamiento, también llamadas competencias blandas. Bajo estas consideraciones, el presente proyecto propone el desarrollo de un plan estratégico para el fortalecimiento de las competencias blandas en el currículo de los programas Ingeniería en Software, Ingeniería Informática, Ciencias Computacionales y afines; suministrando estrategias de enseñanza que reúnan las mejores prácticas para impactar los microcurrículos, dada la naturaleza transferible de las competencias blandas entre industrias y ocupaciones, y en función de las necesidades actuales de formación de la fuerza de trabajo para el sector TI. Para cumplir este objetivo, se empleará una metodología de investigación experimental que facilite un proceso de recolección de datos cuantitativos y cualitativos provenientes de la aplicación de una estrategia de intervención de formación en competencias blandas, fruto de la caracterización de marcos de trabajo, lineamientos y estándares, en un grupo experimental; para luego evaluar los resultados de la implementación mediante la comparación de resultados con el grupo de control. Finalmente, se pretende que el plan estratégico que se obtenga con el desarrollo del proyecto se implemente en el programa Ingeniería en Software de la Facultad de Ingeniería del Tecnológico de Antioquia, de forma que se posibilite un impacto del currículo, a propósito de la renovación de registro calificado del programa y el déficit de los resultados obtenidos por los estudiantes en las pruebas estandarizadas Saber Pro en el periodo comprendido entre 2020 y 2022 para las competencias lectura crítica, razonamiento cuantitativo y competencia ciudadana.

Palabras clave: Competencias Blandas, Ingeniería en Informática, FTyCA-UNCA, Ingeniería en Software, GIISTA-TdeA.

CONTEXTO

En el contexto de los mercados laborales globales, se viene presentando un fenómeno

cada vez más creciente relacionado con la divergencia entre la demanda laboral y la oferta del talento humano. Este desajuste o desequilibrio de habilidades (skills mismatch) se encuentra estrechamente relacionado con el incremento de los niveles de desocupación y la afectación de indicadores como la productividad, la competitividad y la empleabilidad. Así mismo, la carencia de postulantes provistos de las habilidades necesarias para cubrir las vacantes de personal requeridas por el sector productivo favorece las principales problemáticas sociales y económicas que experimentan actualmente los países y pone de manifiesto las profundas brechas de habilidades existentes (Cárdenas Rubio, 2020; Gontero y Novella, 2021; Riomaña y Ríos, 2022). En síntesis, se observa un distanciamiento irrefutable entre el profesional que viene formando la Educación Superior y el talento humano que el mercado laboral demanda en un contexto caracterizado por el cambio constante.

En la línea de la problemática descrita anteriormente, el estudio realizado por la empresa Manpower (2023) líder mundial en soluciones de contratación de talento para puestos permanentes y temporales, analizó los resultados de las entrevistas aplicadas a cerca de 39.000 empleadores en 41 países, encontrando que en promedio el 77% de ellos experimentó dificultades para encontrar el talento humano calificado en lo que va corrido de 2023. Este resultado, comparado con el año inmediatamente anterior, indica el aumento de dos puntos porcentuales y casi el doble de dificultad para encontrar personal adecuado a las necesidades de las compañías con respecto a 2015. En relación con los países de América Latina y el Caribe que participaron en el estudio, los empresarios de Argentina, Brasil, Costa Rica, Guatemala, México, Panamá y Perú; coinciden en la opinión sobre la escasez de personal con las habilidades necesarias para desempeñar los cargos. En estos países, el porcentaje promedio de dificultad para hallar personal con las habilidades requeridas para satisfacer las convocatorias laborales se situó en el 72%. En el caso específico de Colombia, alrededor del 64% de los empresarios encuestados reportó dificultad para cubrir puestos vacantes debido a la escasez de habilidades de los postulantes.

1. INTRODUCCIÓN

La revisión de literatura presenta variadas razones que ocasionan los desequilibrios entre la demanda y la oferta de habilidades que ocurren actualmente en el contexto laboral global. Los estudios elaborados por el Foro Económico Mundial (2018) y la Organización para la Cooperación y el Desarrollo Económicos (OCDE, 2019), destacan el cambio constante y profundo de la tecnología como uno de los impulsores de los desajustes en las habilidades. Por su parte, Gontero y Novella (2021) mencionan la falta de articulación y coordinación entre los sistemas educativos y los mercados laborales para integrar esfuerzos destinados a la planificación de la fuerza laboral, lo que se traduce en una formación de habilidades que difiere de las exigencias del mercado laboral. En cuanto a las políticas de competencias, la OCDE (2019) señala reformas que se encuentran a medio camino entre los sectores educativo, industrial y político; además de la ausencia de coordinación y colaboración entre los ministerios, el gobierno y otras partes interesadas para jalonar iniciativas conducentes a reformar las políticas de competencias.

En la línea de la problemática que se viene describiendo, el marco de desajuste de habilidades existente entre la formación disponible y aquella requerida por el mercado laboral de la industria de las Tecnologías de la Información (TI), también se hace evidente. En este sentido, el estudio sobre la escasez global de talento elaborado por la empresa Manpower (2023) expone que el 78% de los empresarios del sector TI reportó inconvenientes para cubrir vacantes debido a la falta de candidatos en el mercado laboral con las habilidades necesarias para ocupar los cargos; lo que significa el aumento de dos puntos porcentuales con respecto al año 2022. La explicación para la brecha de habilidades existente en la oferta laboral para el sector TI

puede obedecer a la carencia de las habilidades integrales que conjugan aspectos técnicos propios de las ocupaciones del área disciplinar y aspectos generales relacionados con habilidades sociales y de comportamiento.

La discusión sobre las habilidades que se deben formar en los programas académicos relacionados con las TI debe brindar respuesta al proceso de transformación que impacta directamente los mercados laborales. Al respecto, Deloitte (2017) empresa de consultoría de talla mundial, afirma que las habilidades técnicas sólo constituyen una porción de los requisitos para los empleados, y que éstas deben complementarse con atributos personales para lograr el éxito de la fuerza de trabajo en el mercado laboral actual. Por otro lado, el estudio elaborado por el Foro Económico Mundial (2020) argumenta que el futuro del trabajo para el año 2025 se sustenta en una lista de diez habilidades, de las cuales seis constituyen atributos personales y de comportamiento entre los cuales se destacan pensamiento analítico e innovación, aprendizaje activo y estrategias de aprendizaje, resolución de problemas complejos, pensamiento crítico, creatividad, originalidad e iniciativa; liderazgo, influencia social y resiliencia y tolerancia al estrés y flexibilidad.

El escenario descrito anteriormente, plantea una serie de retos que conciernen a la Educación Superior, en tanto que debe velar por un proceso permanente que posibilite el desarrollo integral de las potencialidades de los estudiantes para responder a las necesidades originadas por las nuevas lógicas culturales, económicas, sociales y tecnológicas. La formación de las habilidades requeridas por la fuerza laboral que ocupará cargos en la industria TI es una prioridad para reducir las actuales brechas que impiden la ocupación de los cargos y en este sentido, el Ministerio del Trabajo (2022) resalta la necesidad de alinear las competencias que se imparten en el sistema educativo con los requerimientos del sector productivo, de tal suerte que se favorezca la productividad y la competitividad de las empresas y el acceso a mejores oportunidades de empleo de las personas; aportando de esta manera al desarrollo económico y el progreso social del país.

Por otra parte, los estudios realizados por los principales organismos de carácter multilateral y otros agentes interesados en el estudio de las tendencias que permean la educación indican que las agendas de la Educación Superior deben priorizar la formación de habilidades personales y de comportamiento sin importar el área de conocimiento propio de las ocupaciones. Sobre este punto, la consultora internacional Deloitte (2020) resalta el impulso de las capacidades personales y de comportamiento para ofrecer valor empresarial y adaptar las habilidades técnicas a medida que las necesidades de la fuerza laboral se transforman. En esta misma línea, el Foro Económico Mundial (2018) explica que, a pesar de las fuerzas disruptivas tecnológicas, las personas que tendrán éxito en la economía del futuro serán aquellas que puedan complementar el trabajo realizado por las tecnologías mecánicas o algorítmicas y trabajar con las máquinas.

Las iniciativas que buscan brindar propuestas de solución a las problemáticas vigentes de la Educación Superior, actualizar los currículos en relación con las mejores prácticas para la profundización de la formación integral y procurar la oferta de una capacitación de calidad para los estudiantes; son cada vez más oportunas y necesarias en un mundo fuertemente influenciado por la tecnología, la globalización y los cambios demográficos. En términos de impactos, los proyectos que buscan preparar adecuadamente la fuerza laboral para que responda a las demandas del sector productivo, pueden contribuir decisivamente con los indicadores de productividad, competitividad y empleabilidad; a la vez que constituyen referentes aplicables en el mundo para lograr que los profesionales cumplan las funciones profesionales, investigativas y de servicio que requiere la sociedad.

Teniendo en cuenta los argumentos aludidos anteriormente, la presente propuesta de investigación busca aportar al conocimiento institucional del Tecnológico de Antioquia sobre las mejores prácticas para formar

habilidades blandas en los estudiantes del programa Ingeniería en Software, de forma que se posibilite un impacto del currículo, a propósito de la renovación de registro calificado del programa y el déficit de los resultados obtenidos por los estudiantes en las pruebas estandarizadas Saber Pro en el periodo comprendido entre 2020 y 2022 para las competencias lectura crítica, razonamiento cuantitativo y competencia ciudadana. Igualmente, el proyecto puede impactar positivamente las acciones del área de prácticas interinstitucionales del programa Ingeniería en Software porque pretende articular esfuerzos con los representantes del sector productivo y los egresados, y dejar capacidad instalada para monitorear permanentemente las necesidades de formación requeridas por el mercado laboral. Adicionalmente, se espera contribuir con la función misional del área de extensión institucional de manera que se puedan proponer programas de formación continuada que permitan la preparación de los estudiantes y egresados a lo largo de la vida. Finalmente, dada la naturaleza transferible de las competencias blandas entre industrias y ocupaciones, los resultados del proyecto pueden aplicarse en otras facultades del Tecnológico de Antioquia, en otras universidades nacionales e internacionales y tomarse como referencia en las asociaciones de los programas de ingeniería como ACOFI para replicarse.

El interrogante de investigación que se pretende responder con el desarrollo del presente proyecto de investigación es ¿Cómo integrar en el currículo de los programas académicos Ingeniería en Software, Ingeniería Informática, Ciencias Computacionales y afines estrategias de enseñanza que permitan fortalecer las competencias blandas?

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

La propuesta de investigación se enmarca en un enfoque mixto o de complementariedad, ya que busca integrar y combinar tanto métodos cuantitativos como cualitativos para obtener una comprensión más completa y profunda del fenómeno de interés.

Utilizando técnicas e instrumentos del enfoque cuantitativo, mediante la recopilación de datos numéricos a través de encuestas o cuestionarios, se obtendrá información sobre variables específicas relacionadas con el desarrollo de competencias blandas en estudiantes de Ingeniería en Software, Informática, Ciencias Computacionales y programas afines. Las encuestas se aplicarán a una muestra de estudiantes en diferentes etapas de su formación académica, lo que implica el análisis estadístico de los datos recolectados para identificar patrones, tendencias y correlaciones entre las diferentes estrategias de enseñanza y el desarrollo de competencias blandas; así como la evaluación de la efectividad de las estrategias de enseñanza en función de los resultados obtenidos por los estudiantes en las competencias blandas intervenidas.

Complementariamente, se utilizarán técnicas e instrumentos del enfoque cualitativo mediante la recopilación de datos a través de entrevistas individuales o grupos focales. Este enfoque permitirá obtener información en profundidad sobre la percepción de los estudiantes y profesores sobre las estrategias de enseñanza utilizadas y su impacto en el desarrollo de competencias blandas. Durante las entrevistas o grupos focales, se explorarán experiencias, opiniones, percepciones y sugerencias adicionales relacionadas con la formación en competencias blandas para los programas académicos seleccionados. Estos datos cualitativos serán analizados mediante técnicas de análisis del discurso (hermenéutica) para identificar patrones emergentes y obtener una comprensión más profunda del fenómeno estudiado.

Finalmente, en la etapa de interpretación y análisis, se integrarán los hallazgos de ambos enfoques (cuantitativo y cualitativo) para obtener una visión completa y holística del tema de investigación. La triangulación de los datos permitirá validar y complementar los resultados obtenidos desde ambos enfoques, lo que enriquecerá la comprensión global del impacto de las estrategias de enseñanza en el

desarrollo de competencias blandas en estudiantes de Ingeniería en Software, Informática, Ciencias Computacionales y afines.

La población de interés en esta investigación está compuesta por los estudiantes que se encuentran matriculados en el programa de Ingeniería en Software del Tecnológico de Antioquia (TdeA) y el programa Ingeniería en Informática de la Facultad de Tecnología y Ciencias Aplicadas (FTyCA) de la Universidad Nacional de Catamarca (UNCA), Argentina. Por su parte la muestra seleccionada para este estudio estará compuesta por aquellos estudiantes que están cursando asignaturas del núcleo formativo de Ingeniería en Software.

3. RESULTADOS OBTENIDOS/ESPERADOS

En la línea de la problemática descrita, el estudio realizado por la empresa Manpower (2023) líder mundial en soluciones de contratación de talento para puestos permanentes y temporales, analizó los resultados de las entrevistas aplicadas a cerca de 39.000 empleadores en 41 países, encontrando que en promedio el 77% de ellos experimentó dificultades para encontrar el talento humano calificado en lo que va corrido de 2023. Este resultado, comparado con el año inmediatamente anterior, indica el aumento de dos puntos porcentuales y casi el doble de dificultad para encontrar personal adecuado a las necesidades de las compañías con respecto a 2015.

En relación con los países de América Latina y el Caribe que participaron en el estudio, los empresarios de Argentina, Brasil, Costa Rica, Guatemala, México, Panamá y Perú; coinciden en la opinión sobre la escasez de personal con las habilidades necesarias para desempeñar los cargos. En estos países, el porcentaje promedio de dificultad para hallar personal con las habilidades requeridas para satisfacer las convocatorias laborales se situó en el 72%. En el caso específico de Colombia, alrededor del 64% de los empresarios encuestados reportó dificultad para cubrir puestos vacantes debido a la escasez de habilidades de los postulantes.

Las cifras presentadas anteriormente, demuestran que América Latina posee una notable brecha entre la demanda y la oferta de habilidades laborales. En esta región, más del 50% de los empresarios encuestados hacen alusión a la escasez de habilidades como la principal dificultad para encontrar talento humano calificado.

Es por ello que con base en los resultados de la presente investigación, se preparán conclusiones sobre las estrategias de enseñanza más efectivas para el fortalecimiento de competencias blandas en los estudiantes de los programas seleccionados. Igualmente, se proporcionarán recomendaciones para la implementación futura de estas estrategias en los currículos educativos de los programas tanto de la carrera Ingeniería en Software del Tecnológico de Antioquía - Colombia como de la Ingeniería Informática de la Facultad de Tecnología de la Universidad Nacional de Catamarca.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo del proyecto está formado por docentes investigadores y alumnos de la Facultad de Tecnología y Ciencias Aplicadas y también docentes-investigadores del Grupo de Investigación en Ingeniería del Software (Grupo GIISTA) de la Institución Universitaria Tecnológico de Antioquia (Medellín)

GIISTA, promueve la investigación científica y aplicada en áreas relacionadas con las ciencias de la computación, mediante proyectos de investigación orientados a cuatro líneas: 1) Ingeniería de software, 2) Apoyo tecnológico para ambientes de aprendizaje, 3) Gestión del conocimiento y 4) Seguridad de la información. El grupo fue registrado en el Departamento Administrativo de Ciencia, Tecnología e Innovación (Colciencias) en el año 2004, y en la actualidad goza de excelente posicionamiento al obtener la categoría A en Colciencias.

Los conocimientos generados y experiencia adquirida mediante este proyecto permitirán luego una ampliación y aplicación de los mismos en estrategias de enseñanza-aprendizaje en los planes de estudio de las

carreras antes mencionadas que permitan fortalecer las competencias blandas de manera que los profesionales egresados cuenten con las habilidades necesarias para satisfacer las convocatorias laborales.

Por otra parte, este proyecto aspira entre otras cosas a la formación de profesionales investigadores.

Los integrantes de este Proyecto de Investigación, tienen previsto los años que dura el mismo

- Participar en eventos científicos regionales, nacionales e internacionales de la especialidad, como congresos, simposios, seminarios y cursos, por ello el programa de capacitación y formación de recursos humanos, contempla las siguientes actividades:
- Participar en cursos de actualización y posgrado en el área de estudio.
- Participar en la dirección, asesoramiento y evaluación de tesinas de grado de la carrera de Ingeniería en Infomática de la Facultad de Tecnología y Ciencias Aplicadas de la UNCA y en la carrera Ingeniería de Software de la Institución Universitaria Tecnológico de Antioquia (Medellín)
- Desarrollo de tesis de posgrado por parte de los docentes integrantes.

En cuanto a la formación específica del equipo de docentes investigadores, de la FTyCA-UNCA, se destaca que los mismos se encuentran abocados a actividades de capacitación y estudios de posgrado, para la elaboración de 2 (dos) tesis correspondiente a la carrera Maestría en Ingeniería del Software de la Universidad Nacional de San Luis y 1 (una) tesis de la carrera Maestría en Informática de la Universidad Santo Tomás de Aquino - Tucumán.

5. BIBLIOGRAFÍA

Cárdenas Rubio, J. (2020). A Web-Based Approach to Measure Skill Mismatches and Skills Profiles for a Developing Country. Obtenido de Editorial Universidad del Rosario: https://editorial.urosario.edu.co/gpd-a-web-based-approach-to-measure-skill-mismatches-and-skills-profiles-for-a-developing-country-the-case-of-colombia.html

CASEL. (2020). What Is the CASEL Framework? Obtenido de Collaborative for Academic, Social, and Emotional Learning (CASEL): https://casel.org/fundamentals-of-sel/what-is-the-casel-framework/

Corporación Andina de Fomento (CAF). (2016). Más habilidades para el trabajo y la vida: los aportes de la familia, la escuela, el entorno y el mundo laboral. Obtenido de SCIOTECA Espacio de Conocimiento Abierto: https://scioteca.caf.com/handle/123456789/936

DeakinCo. (Mayo de 2017). Soft skills for business success. Obtenido de Deloitte: https://www.deloitte.com/content/dam/assets-zone1/au/en/docs/services/economics/deloitte-au-economics-deakin-soft-skills-business-success-170517.pdf?trk=article-ssr-frontend-pulse_x-social-details_comments-action_comment-text

Deloitte. (27 de Enero de 2020). Closing the employability skills gap. Obtenido de Deloitte: https://www.deloitte.com/global/en/our-thinking/insights/topics/talent/technology-and-the-future-of-work/closing-the-employability-skills-gap.html

Gontero, S., & Albornoz, S. (2019). La identificación y anticipación de brechas de habilidades laborales en América Latina: experiencias y lecciones. Obtenido de Repositorio Digital Comisión Económica para América Latina y el Caribe: https://repositorio.cepal.org/discover?query=La%20identificaci%C3%B3n%20y%20anticipaci%C3%B3n%20de%20brechas%20de%20habilidades%20laborales%20en%20Am%C3%A9rica%20Latina

Gontero, S., & Novella, R. (2021). El futuro del trabajo y los desajustes de habilidades en América Latina. Obtenido de Repositorio Digital Comisión Económica para América Latina y el Caribe: https://repositorio.cepal.org/discover?query=La%20identificaci%C3%B3n%20y%20anticipaci%C3%B3n%20de%20brechas%20de%20habilidades%20laborales%20en%20Am%C3%A9rica%20Latina

Heckman, J. J., & Kautz, T. (Mayo de 2012). Hard evidence on soft skills. Labour Economics. Obtenido de https://sci-hub.se/10.1016/j.labeco.2012.05.014#:~:text=Hard%20evidence%20on%20soft%20skills.,Labour%20Economics%2C%2019%284%29%2C%20451%E2%80%93464.doi%3A10.1016%2Fj.labeco.2012.05.014%2010.1016%2Fj.labeco.2012.05.014

OCDE. (2019). Estrategia de Competencias de la OCDE 2019. Obtenido de Organización para la Cooperación y el Desarrollo Económicos (OCDE): https://www.oecd.org/general/searchresults/?q=estrategia%20de%20competencias%20de%20la%20ocde%202019&cx=012432601748511391518:xzeadub0b0a&cof=FORID:11&ie=UTF-8

Ríos, L. A., & Riomaña, O. F. (2020). Organización Internacional del Trabajo. Obtenido de Metodología para identificación y medición de brechas de capital humano: https://www.oitcinterfor.org/metodolog%C3%ADa-identificaci%C3%B3n-medici%C3%B3n-brechas-capital-humano-0

World Economic Forum. (Enero de 2018). Eight Futures of Work Scenarios and their Implications. Obtenido de World Economic Forum: https://www3.weforum.org/docs/WEF_FOW_Eight_Futures.pdf

World Economic Forum. (Octubre de 2020). The Future of Jobs Report. Obtenido de World Economic Forum: https://www.weforum.org/reports/the-future-of-jobs-report-2020/in-full

World Economic Forum. (Octubre de 2020). Top 10 skills of 2025. Obtenido de The Future of Jobs Report 2020: https://www.weforum.org/reports/the-future-of-jobs-report-2020/in-full/infographics-e4e69e4de7

PROMINF: SU IMPACTO INSTITUCIONAL EN LA UNLaR

Elena Nicolasa, Pauli

Departamento Académico de Ciencias Exactas Físicas y Naturales
Universidad Nacional de La Rioja
epauli@unlar.edu.ar

RESUMEN

Esta investigación analizó la implementación del Proyecto de Mejoramiento de la Enseñanza en las carreras de Ingeniería y Licenciatura en Sistemas de Información de la Universidad Nacional de La Rioja, financiado por políticas públicas de calidad educativa. A través de proyectos estratégicos evaluados por la Secretaría de Políticas Universitarias, se firmaron convenios para mejorar la calidad tras la acreditación de las carreras. Dentro de las conclusiones más relevantes, puede mencionarse la coincidencia de la población encuestada al reconocer que estos programas fortalecen las carreras si los resultados se sostienen. Además, la experiencia y los antecedentes de estos programas, muestran que fueron una oportunidad y un respaldo económico para las universidades participantes.

***Palabras Claves*:** PROMINF[1], Políticas Públicas, Acreditación, Calidad Académica, Autonomía.

CONTEXTO

Este trabajo fue un proyecto de investigación denominado “PROMINF: su impacto institucional en la UNLaR”; registrado mediante Expte 4036/2022, código 27PIN/A060 y con Resolución de aprobación Nº001/2024 del Consejo de Investigación Científica y Tecnológica CICYT-UNLaR. Se encuadra con las líneas de investigación relacionadas con la Educación Superior, el mismo fue incorporado y ampliado para el desarrollo de la Tesis de Maestría en Educación Superior (UNLaR), aprobado con Resolución de CD.DACHyE[2] Nº193 y cuya defensa oral se realizó el 22 de noviembre de 2023.

1. INTRODUCCIÒN

Los programas de mejoramiento vinculados al financiamiento público han tenido un impacto significativo en las universidades, especialmente en las carreras de Ingeniería y Licenciatura en Sistemas de Información, las cuales han recibido apoyo financiero para mejorar su calidad y formación tras pasar el proceso de acreditación. Esta investigación buscó generar nuevos conocimientos sobre los efectos de estas políticas públicas en la UNLaR, ya que el ejercicio de estas carreras compromete el interés público, por lo que deben atravesar acreditaciones periódicamente por la CONEAU para asegurar la calidad de formación de sus estudiantes.

Entre los antecedentes puede mencionarse que el 22 de julio de 2022, la SPU lanzó una convocatoria para nuevas carreras, destacando como estas iniciativas, fruto del trabajo conjunto entre la SPU y las universidades, han demostrado éxito y sientan las bases para futuros proyectos en otras áreas del conocimiento, con expectativas de mejorar el ámbito académico y científico [1]. Además, la Ley de Educación Superior Nº 24.521/95 [2] establece que las carreras de grado y posgrado reguladas por el Estado deben ser acreditadas; y en este contexto, en 2011 las carreras de Ingeniería y Licenciatura en Sistemas de Información de la UNLaR fueron acreditadas, identificando necesidades como el incremento de ingresantes, la actualización docente y la mejora en los planes de estudio [3]. Es así que CONEAU otorgó una acreditación intermedia de tres años, condicionada al desarrollo de estrategias de mejoras.

La unidad académica del DACEFyN[3] elaboró un plan de mejora financiado por el PROMINF para cumplir con las recomendaciones de los evaluadores. En este contexto, a través del mencionado programa de financiamiento se asignaron fondos según las debilidades detectadas en la acreditación, pero con un enfoque integral adaptado a las necesidades de cada carrera, y no como un mecanismo competitivo [4]. Sin embargo,

[1] Proyecto de Mejoramiento de la Enseñanza en las carreras de Ingeniería y Licenciatura en Sistemas

[2] Consejo Directivo Departamento Académico de Ciencias Humanas y Educación.

[3] Departamento Académico de Ciencias Exactas Físicas y Naturales.

estas políticas no siempre generan los cambios esperados debido a la autonomía institucional y a las modificaciones en la gestión universitaria, lo que, en algunos casos puede generar problemas en la implementación de los mismos [5]

Pese a que el propósito del PROMINF fue mejorar la calidad académica, pocos estudios evaluaron su impacto real, lo que motivó el objetivo de este trabajo investigación, en el marco de la Maestría en Educación Superior de la UNLaR, el cual fue analizar los cambios académicos e institucionales que se habían presentado en las carreras de Ingeniería y Licenciatura en Sistemas de Información como consecuencia de la implementación del PROMINF, en la UNLaR sede Capital, con la finalidad de identificar el impacto y el modo en que se había procesado la implementación según los lineamientos de la SPU y los alcances establecidos haciendo uso de sus márgenes de autonomía universitaria.

Lineamientos Teóricos

Como lo destacan [6] cuando dicen que las políticas públicas educativas, reflejan una creciente responsabilidad de las universidades las cuales deben rendir cuentas de su desempeño académico y financiero a diversas entidades. Del mismo modo, [7] afirman que "el ciclo de la política finaliza con la evaluación de los resultados que han tenido las acciones emprendidas para eliminar, mitigar o modificar ese problema". Según [8], el financiamiento no sigue un mecanismo competitivo, sino que "busca adaptarse a los requerimientos de cada una de las facultades y carreras, adoptando un carácter integral en el abordaje de la problemática".

Por otro lado, se presentan dentro de este sistema, instituciones fundamentales en la tarea académica y científica en el área de las ingenierías en Argentina; como lo es el Concejo Federal de Decanos de Ingeniería (CONFEDI). Qué a nivel internacional, lideró la creación de Asociación Iberoamericana de Instituciones de Enseñanza de la Ingeniería (ASIBEI) junto con la Universidad Politécnica de Madrid, consolidando un foro permanente para la ingeniería iberoamericana. En los años 90, ante la disparidad en titulaciones y formación, se creó el Libro Azul en 1996, unificando currículas y contenidos mínimos. Con la Ley de Educación Superior Nº 24.521 y la creación de la CONEAU, el CONFEDI promovió estándares de acreditación, presentando el Libro Verde en 2000, permitiendo acreditar todas las carreras de ingeniería en el país. En 2006 y 2008, aprobó las competencias de egreso e ingreso para ingenieros respectivamente. En 2018, aprobó los estándares del Libro Rojo, consolidando su rol en la mejora continua de la enseñanza en ingeniería.

A su vez, otra institución relevante es la Comisión Nacional de Evaluación y Acreditación Universitaria (CONEAU), creada en 1996 bajo el Ministerio de Educación, y aplica las normas del sistema universitario definidas por el Ministerio y el Consejo de Universidades [9]; ya que, en este sentido, para [8] "la política contractual y la evaluación de la calidad son procesos indisociables". La CONEAU, es la encargada de evaluar y acreditar carreras de grado y posgrado, su tarea es esencial para garantizar calidad y seguridad en profesiones reguladas (art. 42 y 43, LES). La acreditación, establecida por el Decreto 499/95, es obligatoria para validar títulos profesionales, con revisiones cada seis años. Según [10], la CONEAU logró un cambio cultural al involucrar activamente a las instituciones universitarias. La comisión puede: a) acreditar por seis años, b) acreditar por tres años con planes de mejora, o c) no acreditar si no se cumplen los estándares requeridos [11].

El PROMINF, lanzado el 15 de noviembre de 2012, tuvo como objetivo mejorar la calidad de las carreras de informática mediante proyectos estratégicos no competitivos que fomentaran equidad, calidad y pertinencia [12]. Este programa permitió a las unidades académicas con carreras acreditadas por la CONEAU a presentar proyectos colaborativos y plurianuales. Sus objetivos incluían mejorar la enseñanza en informática, promover la cooperación interuniversitaria y asegurar la pertinencia en función de las demandas sociales. Cada proyecto debía cumplir con requisitos detallados, como presupuestos, estructuras y un convenio-programa que definía obligaciones y financiamiento.

La Resolución Ministerial Nº2017/08 y la Resolución [13] establecieron los procedimientos para la rendición de cuentas y verificación de los recursos asignados. Las carreras de Ingeniería **y** Licenciatura en Sistemas de Información de la UNLaR fueron acreditadas por la CONEAU en 2012 por tres años, comprometiéndose a mejorar aspectos como la investigación, la formación docente y los planes de estudio [14] [15]. En 2017, la acreditación se renovó por otros tres años según acta Nº 474/2017. El PROMINF facilitó proyectos

estratégicos que optimizaron el uso de recursos públicos para mejorar la calidad educativa [8]. Según [16], este proceso reflejó interpretaciones institucionales que adaptaron la política a la cultura organizacional, promoviendo cambios en la gestión académica.

Metodología

El estudio adoptó un enfoque metodológico mixto con predominancia cualitativa, combinando análisis documental, entrevistas semiestructuradas y encuestas cerradas. Según [17], este enfoque permite integrar tanto datos cuantitativos como cualitativos, siendo complementarios en la producción de conocimiento.

El estudio sobre el PROMINF en la UNLaR (sede Capital) incluyó un análisis de documentos oficiales de la SPU, resoluciones de la CONEAU, actos administrativos del DACEFyN, informes de PROMINF y registros institucionales, lo que permitió reconstruir el proceso de implementación del programa. La recolección de datos primarios incluyó entrevistas semiestructuradas a decanos, directora del PROMINF, coordinadores y docentes, también se realizaron encuestas cerradas a docentes y estudiantes que participaron en diversas actividades. Se utilizó un muestreo de conveniencia, seleccionando actores clave según su accesibilidad y relevancia [17].

Además, se analizaron datos cuantitativos, como las estadísticas del SIU Guaraní y las rendiciones financieras de los proyectos presentados, para identificar avances y desafíos. La triangulación de fuentes, combinando entrevistas, encuestas y análisis documental, validó la información. Este estudio exploratorio y parcialmente descriptivo, sin hipótesis previas, tuvo como objetivo conocer el impacto del PROMINF en la calidad educativa de las carreras de Ingeniería y Licenciatura en Sistemas de Información en la UNLaR, generando bases para futuras investigaciones.

2. LÍNEAS DE INVESTIGACIÓN y DESARROLLO

Este proyecto se desarrolló dentro de las líneas de investigación relacionadas con:

- Educación Superior.
- Políticas Públicas.
- Financiamiento Universitario.
- Programas de mejoramiento de las Carreras de Ingeniería y Licenciatura en Sistemas Información de la UNLaR.
- Acreditación de carreras.

3. RESULTADOS OBTENIDOS

En esta sección se presenta una síntesis de los proyectos de mejoramiento diseñados por el DACEFyN de la UNLaR, los cuales fueron programados para ejecutarse en un plazo de tres años. En este sentido, se encontró que la distribución por actividad a fortalecer corresponde en un 33% al Componente A, 11% al Componente B, 12% al Componente C y 44% al Componente D. Esta distribución porcentual fue realizada por actividades en función de los proyectos presentados y los fondos recibidos desde la SPU.

La conformación de los componentes del programa incluía: Componente A (seguimiento de alumnos, graduados, cursos y proyectos de tutoría). Componente B (actividades universitarias, recursos docentes, organización de congresos). Componente C (viáticos, pasantías, movilidad docente, becas de posgrado, formación continua) y Componente D (mejoramiento de infraestructura, equipamiento, software específico, bibliografía y publicaciones).

La población encuestada, coincidió que los planes presentados fueron actividades planificadas para mejorar la enseñanza y responder a las observaciones de los evaluadores. Se destacó que los fondos específicos permiten a las universidades fortalecer sus áreas débiles. También hubo algunas opiniones críticas, las cuales destacaron la oportunidad de mejorar la convocatoria para incluir a más participantes en el diseño de los proyectos, lo que permitiría constituir una mayor diversidad de perspectivas y enriquecer tanto los enfoques como los resultados.

Respecto a los fondos, provenientes principalmente de la SPU, se destaca que estos fueron enviados en tres tramos entre 2014 y 2017, totalizando $1.321.428. Sin embargo, las demoras en los desembolsos afectaron la ejecución oportuna de los proyectos. Mientras tanto el análisis de la matrícula de ingresantes a las carreras de Ingeniería y Licenciatura en Sistemas de Información de la UNLaR mostró que el interés en estas carreras se mantuvo estable a lo largo de la última década, incluyendo el periodo de pandemia. Este comportamiento se relaciona con el proceso de acreditación iniciado en 2011, el acuerdo con la SPU en 2014 y la ejecución de proyectos a partir de 2015.

Los proyectos abarcaban los principales ejes del plan de desarrollo del PROMINF, estructurados según los componentes financiables y sus objetivos. Sin embargo, algunas actividades no se realizaron debido al desajuste entre la convocatoria y la disponibilidad de recursos que se devaluaron debido a la alta tasa de inflación lo que distorsionó los precios iniciales de la propuesta según los datos de INDEC[4], BCRA[5], FACPCE[6], lo que obligó a reprogramar y reorientar los fondos hacia ítems indispensables.

El Componente A del Convenio Nº 1114/14 entre la SPU y la UNLaR se centró en mejorar la gestión académica mediante la migración de la base de datos del SIU Guaraní y la mejora de plataformas como Moodle, optimizando el seguimiento de estudiantes y promoviendo actividades de retención y tutorías.

El Componente B, se enfocó en el intercambio interinstitucional, con la participación de docentes y estudiantes en congresos y eventos como el WICC y el CACIC, fomentando el intercambio de conocimientos, aunque se destacó la necesidad de mejorar la difusión de estos aprendizajes.

El Componente C, destinado al desarrollo de recursos humanos académicos, enfrentó dificultades debido a la desactualización de los fondos y desajustes temporales, lo que afectó su implementación. A pesar de que uno de los objetivos era incrementar la dedicación docente para fortalecer la investigación y extensión, solo tres docentes lograron mejoras en sus dedicaciones, atribuyéndolo a su experiencia e investigaciones previas, este proceso enfrentó desafíos que resaltan la importancia de fortalecer el apoyo institucional para facilitar el logro de estos objetivos.

Por otro lado, el Componente D se centró en la mejora de infraestructura, renovando laboratorios de informática, construyendo uno avanzado y adquiriendo equipos tecnológicos como PCs de alto rendimiento, proyectores y antenas Wi-Fi, lo que mejoró la conectividad y calidad educativa, beneficiando no solo a las carreras de ingeniería y sistemas, sino también a otras disciplinas de la UNLaR. No obstante, el 22% del presupuesto destinado a bibliografía no se ejecutó debido a desajustes entre los fondos presupuestados y los costos actuales.

Los resultados de la encuesta aplicada a los docentes de las carreras de Ingeniería y Licenciatura en Sistemas de Información se basaron en 38 respuestas de un total de 80 docentes encuestados, el 79% de los docentes pertenecen a las carreras de Ingeniería y Licenciatura en Sistemas de Información. La mayoría ocupa cargos de adjunto, y solo 7 docentes ascendieron en la última década, tres gracias al PROMINF. El 52% participó en proyectos de investigación, y la mayoría tiene una categoría de investigación baja (39% en categoría V). En formación académica, el 37% tiene un magíster y el 10% un doctorado. El 32% requiere laboratorios y aulas con conectividad para sus clases. Aunque el 70% conoce el PROMINF, solo el 8% participó en congresos financiados por el mismo.

Los entrevistados coinciden en la necesidad de la continuación de los programas de mejoramiento para actualizar materiales y ajustar planes de estudio. Sin embargo, opinan que deberían ser más amplios y contar con mayores recursos para abordar debilidades detectadas en evaluaciones. La encuesta revela un compromiso docente con la mejora académica, pero también una participación limitada en programas como el PROMINF.

Conclusión

El PROMINF permitió a la UNLaR realizar importantes reformas en infraestructura, como la renovación de laboratorios y la adquisición de equipos tecnológicos, lo que facilitó una mejor enseñanza en informática. Sin embargo, algunos docentes señalaron la falta de participación en los proyectos, lo que indica que los mecanismos para fomentar un trabajo colaborativo no fueron efectivos. A pesar de la prioridad dada a la formación de recursos humanos, los proyectos que se ejecutaron y previstos en este componente, mostró algunas limitaciones en su implementación. Además, el principal problema fue que los recursos no fueron suficientes debido a la alta inflación en el periodo de ejecución, lo que causó que los costos de los proyectos propuestos no se ajustaran al momento de recibir los fondos, impidiendo la realización efectiva de algunas de las propuestas. Bajo las circunstancias mencionadas, los objetivos iniciales del PROMINF fueron parcialmente

[4] INDEC: Instituto Nacional de Estadística y Censo.
[5] BCRA: Banco Central de la República Argentina.
[6] FACPCE: Federación Argentina de Consejos Profesionales de Ciencias Económicas.

alcanzados, en este aspecto se identificaron problemas administrativos y financieros, como la burocracia en el manejo de los fondos, que afectaron los tiempos de ejecución, además de la distorsión del presupuesto inicial. No obstante, los avances más destacados fueron en la informatización de la gestión, mejoras en conectividad, y equipamiento.

A pesar de los logros, la sostenibilidad de estas mejoras y la participación activa de los docentes deben ser áreas de enfoque para futuras convocatorias. Se recomienda continuar investigando sobre proyectos de mejora de calidad educativa en universidades y evaluar su impacto en diversas disciplinas de la UNLaR. Es trascendental mejorar los procesos de seguimiento para optimizar el uso de recursos y evitar reprogramaciones. Además, se sugiere crear unidades estratégicas de negocio para asegurar la sostenibilidad de las mejoras, promover la elaboración de informes y publicaciones para evaluar proyectos, y agilizar el proceso administrativo mediante un gestor de fondos. También se propone desarrollar aplicaciones tecnológicas para monitorear avances, incluir herramientas tecnológicas en los trabajos finales de grado, e implementar una planificación estratégica que permita identificar oportunidades, amenazas, fortalezas y debilidades internas.

4. FORMACION DE RECURSOS HUMANOS

Las investigaciones previas y los relevamientos de datos de este proyecto de investigación, han permitido ser bases para luego ser desarrollado como tema de tesis de maestría, lográndose la finalización del trabajo de Tesis de Maestría en Educación Superior de la autora. También el proyecto fue integrado por estudiantes avanzados de la carrera Ingeniería y Licenciatura en Sistemas, como así también, docentes y nodocentes que se formaron en este equipo de investigación.

5. BIBLIOGRAFIA

[1] SPU. (2022). Proyecto de mejora de la formación para las carreras de Licenciatura en Biotecnología y Genética https://www.argentina.gob.ar/sites/default/files/2022/09/if-2022-75575148-apn-secpume.pdf

[2] Ministerio de Educación. (1995). Ley 24521/95. Honorable Consejo de la Nación. https://www.argentina.gob.ar/normativa/nacional/ley-24521-25394/texto

[3] CONEAU. (2017). Anexo. https://www.coneau.gob.ar/archivos/anexos/IF-2017-33152458-APN-DAC-CONEAU.pdf

[4] García de Fanelli, A. y Claverie, J. (2013). Políticas públicas para mejorar la calidad de las carreras de ingeniería. Estudio de caso del PROMEI en Argentina. Revista Argentina de Educación Superior, 5(6), 72-96.

[5] Oszlak, O. (2003). Estudio sobre el impacto del Fondo para el Mejoramiento de la Calidad Universitaria (FOMEC). En J. C. Pugliese (Ed.), Políticas de Estado para la Universidad Argentina (pp. 183-188). Ministerio de Educación, Ciencia y Tecnología Secretaria de Políticas Universitarias. http://www.bnm.me.gov.ar/giga1/documentos/EL002743.pdf

[6] Machi, J., Broto, A., (2003). La información como herramientas para la toma de decisiones. En J.C. Pugliese (Ed.), Políticas de Estado para la Universidad Argentina (pp. 196-200).

[7] Clausen, F. y Foglia, C. (2012). Políticas públicas y gestión compartidas - Módulo 4. Diplomatura Fortalecimiento de las Capacidades de Gestión de las Organizaciones Sociales Territoriales, IV. https://www.argentina.gob.ar/sites/default/files/2.4politicas_publicas_y_gestion_compartida.pdf

[8] García de Fanelli, A. M. (2006). Los contratos-programa en las universidades: Lecciones de la comparación internacional. Archivos Analíticos de Políticas Educativas, 14(11). (pp. 1-26) https://www.redalyc.org/pdf/2750/275020543011.pdf

[9] Villanueva, E. (2008). La acreditación en contexto de cambio: El caso de las carreras de ingeniería en la Argentina. Revista da Avaliação da Educação Superior, 13(3), 793-805. https://www.redalyc.org/articulo.oa?id=219114874009

[10] Pugliese, J. C.(Ed.) (2003). Políticas de Estado para la Universidad Argentina. Ministerio de Educación, Ciencia y Tecnología. Secretaria de Políticas Universitarias. http://www.bnm.me.gov.ar/giga1/documentos/EL002743.pdf

[11] CONEAU. Buscador de Carreras https://www.coneau.gob.ar/buscadores/grado/buscar.php

[12] SPU. (2012). Resolución Nº 2290. Presidencia de la Nación - Ministerio de Educación - Secretaría de Políticas Universitarias. Buenos Aires.

[13] SPU. (2010). Resolución Nº 2260. Normas generales de Rendición de Cuentas. http://secat.unicen.edu.ar/wp-content/uploads/2017/04/RSPU-2260-10-Rendici%C3%B3n.pdf

[14] CONEAU. (2012) a). Resolución Nº 475/12. Acreditación con compromisos de mejoramiento la carrera de Licenciatura en Sistemas de Información del Departamento Académico de Ciencias Exactas Físicas y Naturales de la UNLaR. Buenos Aires https://www.coneau.gob.ar/archivos/resoluciones/Res475-12E804082910.pdf

[15] CONEAU. (2012). b) Resolución Nº 476/12. Acreditación con compromisos de mejoramiento la carrera de Ingeniería en Sistemas de información del Departamento Académico de Ciencias Exactas Físicas y Naturales de la UNLaR. Buenos Aires. https://www.coneau.gob.ar/archivos/resoluciones/Res476-12E804083010.pdf

[16] Miranda, E. (2011). Una "caja de herramientas" para el análisis de la trayectoria de la política educativa. La perspectiva de los ciclos de la política. En E. Miranda, L. Aguilar, N. Paciulli Bryan, S. Roitenburd, (Re) Pensar la educación pública (1° ed., pp. 105-126). Facultad de Filosofía y Humanidades UNC. https://rdu.unc.edu.ar/bitstream/handle/11086/546337/Miranda_Una_caja_de_herramientas.pdf?sequence=3

[17] Hernández Sampieri, R., Fernández Collado, C. y Baptista Lucio, P. (2014). Metodología de la investigación (6° ed.). Mc-Graw Hill Education.

Unidades de proyectos de investigación en Sistemas de Información

Bianchini Germán[1], Caymes-Scutari Paola[1,2], Monetti Julio[3], Ontiveros Patricia[4], Rotella Carina[4], Tagarelli Sandra[5]

[1]Laboratorio de Investigación en Cómputo Paralelo/Distribuido, Departamento de Ingeniería en Sistemas de Información. Facultad Regional Mendoza/Universidad Tecnológica Nacional
Rodríguez 273 (M5502AJE) Mendoza, +54 261 5244579
[2]Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)
[3]Laboratorio de Integración de Tecnologías Aplicadas a Propotipos de Software, Departamento de Ingeniería en Sistemas de Información. Universidad Tecnológica Nacional, FRM. Rodriguez 273, (M5502AJE) Mendoza.
[4]Laboratorio de Gobierno Electrónico, Departamento de Ingeniería en Sistemas de Información. Facultad Regional Mendoza/Universidad Tecnológica Nacional
Rodríguez 273 (M5502AJE) Mendoza
[5]Laboratorio de Analítica de Datos, Departamento de Ingeniería en Sistemas de Información. Facultad Regional Mendoza/Universidad Tecnológica Nacional
Rodríguez 273 (M5502AJE) Mendoza

gbianchini@frm.utn.edu.ar, pcaymesscutari@frm.utn.edu.ar, jmonetti@frm.utn.edu.ar, pontiveros@frm.utn.edu.ar, carinarotella@gmail.com, stagarelli@gmail.com,

RESUMEN

Entre los objetivos principales de una universidad se encuentra la generación, desarrollo, difusión y transferencia de conocimiento a través de la investigación científica.

Para lograrlo, es esencial formar individuos y grupos capaces de impactar positivamente en la comunidad académica y la sociedad, abordando desafíos diversos y demostrando un sólido compromiso con la misma. Con esta visión, en abril de 2023, iniciamos un proyecto compuesto por docentes y estudiantes con el objetivo de guiarlos y prepararlos para los diversos aspectos que implica la investigación científica.

Dentro de esta propuesta, se destaca la posibilidad con la que cuentan los participantes de proponer y llevar adelante sus propios proyectos de investigación (denominados Unidades de Proyecto o UP), lo que permite a los investigadores en formación indagar y desarrollar sus propios temas de interés, sin la necesidad de contar con los antecedentes y *expertise* que normalmente requiere la formulación y dirección de un proyecto clásico.

El proyecto globalizador que abarca a todas las UP constituye un marco para la promoción de ideas dentro del proceso de iniciación en la investigación.

Los resultados logrados a lo largo de estos años de proyecto incluyen valiosos aportes, la participación en eventos científicos y, sobre todo, la transferencia de conocimientos tanto a docentes como a estudiantes.

***Palabras clave*:** Iniciación, Alumnos, Docentes, Vocación.

CONTEXTO

El proyecto donde se encuentra enmarcado el presente trabajo, proyecto titulado 'Marco para desarrollar nuevas unidades de proyectos e incorporar docentes y alumnos en actividades de investigación en ISI' (PID TETEUME0008760TC) se encuentra en su tercer año de desarrollo. El mismo se realiza en el ámbito del LICPaD (Laboratorio de Investigación en Cómputo Paralelo /Distribuido), contando con la participación tanto de docentes del propio laboratorio, como de docentes de otros grupos de la institución, como ADA-Lab (Laboratorio de Analítica de Datos), GE-Lab (Laboratorio de Gobierno Electrónico) y LITAPS (Laboratorio de Integración de Tecnologías Aplicadas a Propotipos de Software), todos pertenecientes al Departamento de Ingeniería en Sistemas de Información de la UTN-FRM. En este proyecto se busca aunar la formación teórica y la práctica de investigadores iniciales para que éstos experimenten todas las etapas necesarias que involucra el proceso científico en un ámbito de investigación. Para lograr dicho objetivo se utiliza un trabajo de investigación específico como meta del proceso de aprendizaje, enfocado en las competencias específicas que deben adquirir los graduados de Ingeniería en Sistemas de Información, además de impulsar el pensamiento crítico de docentes y alumnos frente a nuevos retos. En el caso del LICPaD, los participantes de la formación son alumnos y becarios, mientras que en el caso de ADA-Lab, GE-Lab y LITAPS, los participantes en formación son los propios docentes.

1. INTRODUCCIÓN

Mediante estrategias formativas centradas en la realización de investigaciones específicas se pueden abordar los objetivos de formación científica y contribuir al objetivo general de promover vocaciones científicas como apuesta clave para construir una sociedad del conocimiento sólida [1, 2, 3, 4, 5]. De esta manera, analizando la literatura educativa acerca de la enseñanza a través de la investigación y de las competencias científicas, diversos estudios resaltan los diferentes aspectos que intervienen en la motivación general de los individuos [6]. Es importante crear un clima de motivación hacia la tarea que debe realizarse, lo que supone desarrollar nuevas estrategias para la metodología investigadora. La utilización de actividades de indagación e investigación como estrategia didáctica permite favorecer aprendizajes significativos y promover las competencias potenciando el desarrollo de sus aptitudes científicas [7, 8, 9].

El Marco para desarrollar Unidades de Proyecto es, además de un elemento pedagógico, un proyecto de investigación propiamente dicho, liderado por científicos con experiencia en el campo. Por tanto, ha de concebirse como tal, con unos objetivos, con información previa sobre el área en la que se va a desarrollar cada UP, con tareas y materiales adecuados, y un cronograma factible. Todo esto ejerce un efecto positivo en la consecución y en la obtención de resultados que no han de ser meramente científicos y/o ponderables. La verdadera meta es la formación del alumno o docente en el mundo de la ciencia y, por qué no, en lo social, ya que le permitirá socializar en un entorno nuevo para el que originalmente no estaba preparado.

La necesidad de fortalecer la formación de investigadores se refleja también en los procesos de acreditación de las carreras universitarias. La evaluación de estas carreras incluye como aspecto

fundamental la producción de proyectos de investigación y los aportes del cuerpo docente y, en la medida de lo posible, del alumnado.
La UTN ofrece recursos para la formación de investigadores [10], tales como cursos que brindan información sobre buenas prácticas en la vinculación tecnológica, gestión de la ciencia, tecnología e innovación, e inglés para la publicación de artículos científicos, entre otros.
Sin embargo, se observa una brecha y un desconocimiento por parte de los potenciales investigadores, producto de la desinformación y la falta de instrumentos específicos para profesionalizar y financiar la investigación. Esto limita el acceso a los recursos y fuentes de información disponibles.
A pesar de estas dificultades presupuestarias y para que no se trate sólo de una actividad vocacional, el presente proyecto busca allanar el camino a aquellos interesados en desarrollar sus propios proyectos de investigación e insertarse en el ámbito científico, brindando apoyo y fomentando el desarrollo de las investigaciones en la medida que los recursos lo permitan.

2. LÍNEAS DE INVESTIGACIÓN y DESARROLLO

Se ha trabajado en diferentes Unidades de Proyecto, en donde cada Unidad se traduce en un proyecto individual o grupal, según sea el caso, en el que los investigadores orientan sus actividades de acuerdo a sus conocimientos y motivaciones.
Las UP específicas llevadas a cabo son las siguientes:
_Uno de los equipos de trabajo, conformado por cinco estudiantes de segundo año de la carrera, ha trabajado en una UP relacionada con el reconocimiento de imágenes (números escritos a mano) mediante algoritmos genéticos.
_Otro equipo de estudiantes, en este caso compuesto por tres estudiantes avanzados, ha trabajado sobre la resolución del problema TSP (*Travelling Salesman Problem*) mediante la aplicación del modelo ACO (*Ant Colony Optimization*). A su vez, este equipo se dividió en dos subgrupos, donde uno se concentró en las mejoras del modelo y otro en la aplicación de GPU para la reducción del tiempo de ejecución mediante paralelismo.
_Una UP individual de un alumno avanzado de Electrónica se ha enfocado en el estudio del paradigma Paralelo/Distribuido.
_Otro caso de una UP de un único estudiante inicial se ha orientado al estudio de las posibilidades del uso de GPUs en el procesamiento paralelo.
_En relación a los docentes, uno de los equipos, conformado por cuatro docentes, orientó su actividad en la aplicación de IA para servicios de internación domiciliaria, y también en el uso de tecnología Blockchain para la emisión de certificados de extensión universitaria.
-Finalmente, un equipo individual se ha concentrado en el uso de AC (Autómatas Celulares) y su aplicación mediante un framework de simulación, así como también a la programación basada en agentes.
En todos los casos, los equipos de trabajo han sido acompañados y asistidos por los directores del LICPaD, quienes conducen y coordinan el Proyecto Marco global.

3. RESULTADOS OBTENIDOS

Tras dos años de proyecto (actualmente se transita el tercero), hemos conseguido algunos resultados interesantes situados

en dos niveles distintos. El primer nivel tiene que ver con los objetivos específicos que persigue el proyecto global, esto es, promover la vocación científica de alumnos y docentes y a la vez asistirlos en la inserción en dicho medio. En este mismo plano también se ha trabajado en la mejora de las habilidades en el uso de herramientas para realizar investigación, la organización y estimación de tiempos y cronogramas de trabajo, y en la comunicación de resultados mediante la práctica del proceso de redacción, corrección, revisión, publicación y presentación de los aportes logrados. Esta clase de resultado se ha alcanzado de acuerdo a las capacidades iniciales de los investigadores en formación, dado que algunas UP cuentan con participantes alumnos que se encuentran en diferentes años de la carrera, y en otros casos las unidades se conforman con docentes quienes en algunos casos no poseían experiencia de las tareas que involucra un proyecto de investigación, aunque en otros casos se trata de docentes que han cursado una maestría o se encuentran realizando algún doctorado.

En el segundo nivel aparecen los resultados específicos de las distintas UP. Seguidamente, se describen los resultados específicos de cada una. En el caso de la UP de Reconocimiento de Imágenes, grupo conformado por estudiantes de segundo año de la Ing. en Sistemas de Información, el logro que han obtenido ha sido el desarrollo de una primera versión del algoritmo a la que han entrenado mediante una base de datos para el reconocimiento de números manuscritos.

La segunda UP, conformada por estudiantes avanzados de la Ing. en Sistemas, trabajó sobre la versión inicial de la resolución del problema TSP a través de la aplicación del modelo ACO. Sobre dicha implementación se realizaron múltiples experimentos con distintos números de ciudades para evaluar la respuesta, notando una baja escalabilidad. Por esta razón, la UP se organizó en dos equipos, donde uno se concentró en las mejoras del modelo, aplicando recorridos mediante las k ciudades más cercanas, y otro equipo se concentró en la aplicación de GPU para la mejora de los tiempos de respuesta.

Las dos UP individuales de alumnos realizaron un relevamiento del Estado del Arte en sus respectivas temáticas e hicieron pruebas iniciales (por ejemplo, la implementación del cálculo de la integral mediante la suma de Riemann para la comparación del rendimiento secuencial vs el paralelo).

Una de las UP conformada por docentes, ha indagado en diversas temáticas en base a sus intereses (aplicación de IA para servicios de internación domiciliaria; uso de tecnología Blockchain para la emisión de certificados de extensión universitaria; etc.).

Finalmente, el docente investigador de la UP individual ha enfocado la labor en el uso de autómatas celulares y su aplicación mediante un framework de simulación.

En todos los casos, los resultados preliminares se han presentado en el marco del XII ENIDI realizado en Los Reyunos (Mendoza), en el XII CONAIISI (Catamarca), en el CADI CAEDI 2024 (San Juan) y en ediciones previas del WICC.

4. FORMACIÓN DE RECURSOS HUMANOS

Durante los dos primeros años de proyecto, han participado del mismo cinco docentes y diez estudiantes, además de los dos investigadores formados que lo conducen y coordinan. Algunos de los participantes alumnos son becarios de

grado y cuentan con becas de investigación (BIS o I+D), que, si bien son de montos bajos, les brindan formalidad y una mínima ayuda para dedicar algunas horas semanales a la investigación, aunque no hay nivel de comparación con los elevados ingresos que pueden percibir en el ámbito privado. No obstante, esto les ha permitido cumplir con los objetivos del proyecto y generar antecedentes para su futuro profesional.

Algunos docentes en formación tienen una dedicación simple para tareas de investigación, pero requieren guía para abordar sus proyectos y obtener financiamiento propio a largo plazo.

5. BIBLIOGRAFÍA

[1] National Research Council. (2013). Next generation science standards. Board on Science Education, Center for Education. Division of Behavioral and Social Sciences and Education. Washington, DC: National Academies Press.

[2] Aldana, M. (2012). ¿Qué le falta a la ciencia en México? Revista Temas, 69, pp. 26-30.

[3] Sanz Menéndez, L., Cruz Castro, L. (2010). Análisis sobre ciencia e innovación en España. España: Instituto de Políticas y Bienes Públicos del CSIC.

[4] CONACYT. (2014). Programa Especial de Ciencia, Tecnología e Innovación 2014-2018. México: Gobierno de la República.

[5] Escalante, H., Corrales, R., Barahona, M. (2011). La investigación científica en los estudiantes de grado y posgrado. Revista Portal de la Ciencia, UNAH. Nro.11. pp.19-25.

[6] Ryan, R. M. & Deci, E.L. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and. New Directions. Contemporary Educational Psychology 25, 54-67.

[7] Caamaño, A. (2012). La investigación escolar es la actividad que mejor integra el aprendizaje de los diferentes procedimientos científicos. En Pedrinaci, E., Pro, A., Caamaño, A. y Cañal, P. (coords.). 11 ideas claves. El desarrollo de la competencia científica, 127-146. Graó.

[8] Couso, D. (2012). "De la moda de "aprender indagando" a la indagación para modelizar: una reflexión crítica". Conferencia plenaria inaugural de los XXVI Encuentros de Didáctica de las Ciencias Experimentales de APICE. Huelva. (fecha de consulta: febrero de 2025. http://uhu.es/26edce/actas/docs/conferencias/pdf/26ENCUENTRO_DCE-ConferenciaPlenariaInaugural.pdf

[9] Lupión, T. y Martín, C. (2016). Scientific school research: in-service teachers assessment of educational contents and strategies. In J. Lavonen, K. Juuti, J. Lampiselkä, A. Uittoy and K. Hahl (eds.) Science Education Research: Engaging learners for a sustainable future, 1466-1472.Helsinki: ESERA.

[10] Cursos PROFORVIN. https://www.utn.edu.ar/es/secretaria-academica/proforvin (fecha de consulta: febrero de 2025)

· ISS ·

Innovación en Sistemas de Software

Avances en la Línea de investigación: Motor de aumentación de prompt modelos de IA generativos

Dr. Martin Becerra [1], **Dr. Jorge Ierache** [1,2], **Mg. Pablo Soligo** [2], **Ing. Diego Sanz** [1]
[1] **Universidad Nacional de La Matanza, DIIT, Grupo de Investigación de Realidad Aumentada e Inteligencia Artificial Aplicada (RAIA)**
[2] **Universidad Nacional de La Matanza, DIIT, Grupo de Investigación y Desarrollo de Software Aeroespacial (GIDSA)**
Florencio Varela 1903, La Matanza, Buenos Aires, Argentina
{mabecerra, jierache}@unlam.edu.ar

Resumen

Se presentan en este trabajo los avances de la línea de investigación dedicada a explorar el desarrollo de framework para la asistencia, creación de prompts e integración de datos para modelos IA generativa. El framework busca facilitar la aplicación de técnicas de aumentación de prompts en aplicaciones potenciadas por modelos de lenguaje (LLMs).

Palabras clave: Modelos generativos, Grafos de conocimiento (Knowledge graphs), Frameworks, Modelos de Lenguajes

Contexto

La investigación presentada es desarrollada por el Grupo de Investigación de Realidad Aumentada e Inteligencia Artificial Aplicada del Departamento de Ingeniería e Investigaciones Tecnológicas de la Universidad Nacional de La Matanza en el marco del proyecto PROINCE C256: Framework para explotación de modelos de IA generativos multimodales.

Introducción

En la actualidad se han logrado varios avances en modelos de lenguaje de Inteligencia artificial que mejoraron los modelos GPT 3 y 4[1] , ChatGPT[2], Llama2[3] y Gemini[4] aplicados a tareas de propósito general multimodales (Ejemplo: texto, imágenes, video, y razonamiento estadístico). Deepseek logró con sus modelos V3[5] y R1[6] que el costo de entrenamiento se reduzca considerablemente lo que permite tener capacidades similares a sistemas GPT y modelos grandes con modelos más pequeños. Esto generó un gran impacto en la industria y en la investigación para poder replicar resultados obtenidos por Deepseek[7].

La evolución de IA generativa, muestra una gran capacidad para modelar el lenguaje, sin embargo presentan inconvenientes, una de ellos que por generar palabras estadísticamente estos modelos generativos dependen fuertemente del prompt inicial ofrecido por el usuario para que se obtengan los resultados deseados. Otro inconveniente, derivado del anterior, es que estos modelos sufren de alucinaciones[8], en otras palabras el modelo devuelve datos que no corresponden al prompt enviado por el usuario. Para abordar el primer inconveniente surgió la disciplina ingeniería de prompts[9] que busca estudiar los diferentes patrones de prompts y técnicas para incorporar conocimiento externo que permiten alcanzar mejores resultados. Para el segundo, se encuentra una técnica propuesta por facebook llamada Retrieval augmented generation (RAG)[10], que permite integrar hechos fácticos / datos desde base de datos y/o Grafos de conocimiento (Knowledge graphs)[11] en la inferencia del modelo generativo (Large Language model, LLM) [12], esta estrategia tiene como desventaja que es compleja de llevar a cabo correctamente.

Por lo tanto, bajo la creciente tendencia de creación de nuevas técnicas para generar prompts tales como chain of thoughts[13], a tree of thoughts[14] y Multimodal chain-of-

thought[15] e integrar hechos fácticos mediante para extender el conocimiento de un LLM es cada vez más complejo mediante técnicas tales como RAG [10] y GraphRAG[16] integrar estos enfoques en aplicaciones para ingenieros sin conocimientos técnicos, en particular en Procesamiento de Lenguaje Natural (NLP) y modelado de conocimiento.

Líneas de Investigación, Desarrollo e Innovación

La línea de investigación y desarrollo se lleva adelante bajo el marco del proyecto PROINCE C-256: Motor de aumentación de prompt modelos de IA generativos.

El objetivo de esta línea de investigación aplicada es continuar la línea propuesta en [17] es el desarrollo de un motor para reducir la complejidad de desarrollo de aplicaciones potenciadas por LLMs. Se propone que el framework permita de manera preliminar abstraer y reducir complejidad de algunos de estos aspectos listados a continuación:

- Creación de prompts.
- Selección de estrategias de aumentación de prompts.
- Incorporación de información de dominio externos del framework mediante la incorporación de bases de datos y/o Knowledge graphs.

El desarrollo del Framework es potencialmente aplicable en diversos dominios, las pruebas preliminarmente se enfocarán en la selección de un dominio particular (educación, salud, industria y gastronomía).

Resultados y Objetivos

Los resultados obtenidos contribuyen con el objetivos de obtener un motor que permita aumentar el prompt de un usuario, con el desarrollo de un framework; el motor recibirá como entrada el prompt del usuario y el del sistema, selecciona la aumentación del prompt y agrega información de dominio en forma de documentos / grafos de conocimiento que son externos al framework. como se indica en la figura 1.



Figura 1 Esquema conceptual Framework para la aumentación de prompts para modelos generativos de IA.

El objetivo del motor es aumentar el prompt. Para ello, se realizan un conjunto de transformaciones. En la figura 2 se puede observar la secuencia de pasos (En inglés Pipeline) que se busca aplicar al prompt del usuario. En el paso 1 , se busca determinar el dominio de la tarea que quiere realizar el usuario. Una vez detectado, se procede al paso 2 Selección de estrategia de aumentación en el que el motor decide qué estrategia de aumentación utilizar. En el paso 3, se creará un contexto para poder agregar información adicional que no está presente en el texto del prompt a partir de fuentes externas de datos. Por último, en el paso 4 se alimenta el prompt aumentado al LLM elegido (LLama2, Gemma o GPT).

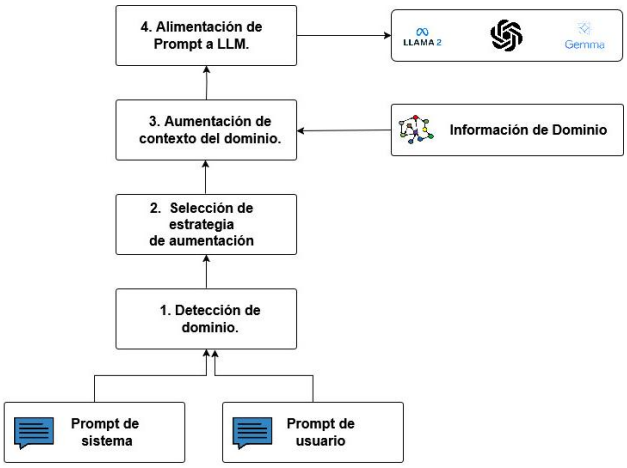


Figura 2 Esquema conceptual del pipeline que realiza el motor de aumentación de prompts.

De esta manera, se busca que el usuario obtenga los resultados deseados mediante la secuencia de transformaciones mencionadas. Un detalle a destacar es que se busca que el framework sea independiente del modelo generativo utilizado, ofreciendo la posibilidad de integrar cualquier modelo generativo (LLama2, GPT o Gemma).

Formación de Recursos Humanos

El grupo de investigación está conformado por tres investigadores formados y un investigador en formación, trabajando en el área de IA.

Referencias

[1] GPT-4 Technical Report. Disponible en https://arxiv.org/abs/2303.08774. Accedido el 5/2/2024

[2] OpenIA, ChatGPT. Disponible en https://openai.com/blog/chatgpt. Accedido el 12/2/2025.

[3] Facebook, LLama 2. Disponible en https://llama.meta.com/. Accedido el 12/2/2025.

[4] Google, Gemini. Disponible en https://deepmind.google/technologies/gemini/#introduction. Accedido el 12/2/2025.

[5] Liang, W., et al. (2024). DeepSeek-V3 Technical Report. Disponible en https://arxiv.org/abs/2412.19437. Accedido el 12/2/2025.

[6] Liang, W., et al. (2025). DeepSeek-R1: Incentivizing reasoning capability in LLMs via reinforcement learning . Disponible en https://arxiv.org/abs/2501.12948 . Accedido el 12/2/2025.

[7] Hugging Face Team. (2024). Open-R1: A fully open reproduction of DeepSeek-R1. Disponible en https://huggingface.co/blog/open-r1 Accedido el 12/2/2025.

[8] Wikipedia, Hallucination. Disponible en https://en.wikipedia.org/wiki/Hallucination_(artificial_intelligence). Accedido el 5/2/2024.
Open IA, Prompt engineering. Disponible en https://platform.openai.com/docs/guides/prompt-engineering. Accedido el 12/2/2025.

[9] IBM. (2024). Prompt engineering. Disponible en https://www.ibm.com/es-es/think/topics/prompt-engineering. Accedido el 12/2/2025.

[10] Facebook, Retrieval Augmented Generation. Disponible en https://ai.meta.com/blog/retrieval-augmented-generation-streamlining-the-creation-of-intelligent-natural-language-processing-models/. Accedido el 12/2/2025.

[11] IBM, Knowledge graphs. Disponible en https://www.ibm.com/topics/knowledge-graph. Accedido el 12/2/2025.

[12] Nvidia. Large Language Models Explained disponible en https://www.nvidia.com/en-us/glossary/large-language-models/. Accedido el 12/2/2025.

[13] Wei, J. et. al. (2023). Chain-of-thought prompting elicits reasoning in large language models Disponible en https://arxiv.org/abs/2201.11903. Accedido el 12/2/2025.

[14] Yao, S. et. al. (2023). Tree of thoughts: Deliberate problem solving with large language models. Disponible en https://arxiv.org/abs/2305.10601. Accedido el 12/2/2025.

[15] Zhang, A. et. al. (2023). Multimodal chain-of-thought reasoning in language models disponible en https://arxiv.org/abs/2302.00923. Accedido el 12/2/2025.

[16] Edge, D., et al. (2024). From Local to Global: A Graph RAG Approach to Query-Focused Summarization. Disponible en https://arxiv.org/abs/2404.16130. Accedido el 12/2/2025.

[17] Becerra, et. al. (2024). Línea de investigación: Framework para explotación de modelos de IA generativos. Disponible en https://wicc2024.unp.edu.ar/images/libros/LibroDeActasWICC2024.pdf. Accedido el 12/2/2025.

El valor del uso de los datos aplicado a la estrategia alimentaria

Ing. Franco Quintero [1,7] C.C Marcelo Taruschio [2,7]
Dr. Edgar Gonzalo Cossio Franco[3,8] AS David Esteban Clai[4,7], AS Analía Forte[5,7,] AS Nicolas Campetti[6,7]

[7]Facultad de Ciencias Exactas e Ingeniería. Universidad Católica de La Plata.
[8]Universidad del Valle de Atemajac, UNIVA, México
[1] francopablo.quintero@ucalpvirtual.edu.ar [2] marcelo.taruschio@ucalp.edu.ar
[3]edgar.cossio@univa.mx [4] foxblue70@gmail.com [5] analia_fff@hotmail.com
[6] nicolas_campetti@hotmail.com

Resumen

En el marco del proyecto "Estrategia de distribución holística de alimentos como herramienta de gestión. Su aplicación en el Banco Alimentario", esta investigación tiene como objetivo delinear el estado de situación y la problemática actual del Banco Alimentario de la Ciudad de La Plata. En ese sentido se identifican varias áreas de estudio cruciales, tales como, gestión de proveedores, ingreso y egreso de mercadería, gestión de inventarios y la distribución de mercadería para que, a partir del análisis y diseño se pueda crear una herramienta de software (modulo) que permita mejorar el resultado de distribución de alimentos a los beneficiarios (comedores) del Banco Alimentario, el objetivo es reducir el costo de distribución de alimentos.

Se presentan avances sobre la investigación realizada. El propósito general, en una primera etapa es identificar y especificar los requerimientos para el desarrollo de una herramienta de software que permita reducir o mitigar los costos asociados a la distribución de alimentos y mercaderías a los diferentes beneficiarios (comedores), identificando y priorizando aquellos con mayores necesidades.



De manera adicional se generará como producto entregable un documento detallado que describe la estrategia de distribución holística, incluyendo:

- Visión general de la estrategia.
- Análisis de la situación actual.
- Objetivos y principios de la estrategia.
- Estrategias específicas para optimizar la logística, la gestión de inventario, utilizando herramientas de IA.
- Plan de acción detallado.
- Indicadores clave de rendimiento (KPIs).

Palabras clave:
Gestión de proyectos de software. Logística. IA aplicado a software de logística. Análisis de datos aplicado a la estrategia alimentaria.

Contexto

La investigación surge como una línea de trabajo colaborativa con la organización Civil Banco Alimentario de la ciudad de La Plata. Ante el creciente avance de la necesidad de satisfacer la distribución y recolección de los alimentos con la mayor inmediatez, esta investigación se enfoca en enfrentar los desafíos asociados a los costos de adquisición y distribución.

Si bien el resultado del trabajo está enfocado en la distribución de los alimentos, el análisis y diseño se realizó sobre todo el estado de situación (íntegro) del Banco Alimentario.

A continuación, la metodología y pasos aplicados:

• Recopilación de información sobre las mejores prácticas en distribución holística de alimentos, análisis del contexto del Banco Alimentario y sus necesidades específicas.

• Evaluación de la capacidad actual de distribución, identificación de brechas y oportunidades de mejora.

• Desarrollo de un plan de acción detallado que incluya estrategias para optimizar la logística, la gestión de inventario, utilizando herramientas de IA y explotación de datos obtenidos del relevamiento.

• Puesta en marcha del plan de acción, seguimiento y evaluación de resultados.



Nota: El comedor beneficiario es el único responsable de buscar la mercadería asignada por el Banco Alimentario.

Introducción

En el contexto global actual, la inseguridad alimentaria persiste como un problema grave, mientras grandes cantidades de alimentos comestibles se desperdician anualmente. Organizaciones como los bancos de alimentos han surgido para enfrentar este desafío, aunque a menudo lo hacen sin un enfoque estratégico y holístico. Este proyecto propone optimizar la gestión de recursos, reducir el desperdicio de alimentos y mejorar la ayuda a la comunidad, potenciando la sostenibilidad a largo plazo de la organización desde la óptica de la distribución de alimentos.

Siguiendo la naturaleza del desarrollo de un producto de software, la estrategia se forja a partir de la identificación de un mapa de impacto, que sienta las bases de las necesidades de los requerimientos y expectativas de los usuarios. A partir de las necesidades se definirá un roadmap de historias de usuarios hacia el logro del producto deseado. Con las historias de usuarios identificadas se creará el “release plan” junto con su MVP (Mínimo Producto Viable) como punto de partida y se empleará una metodología Agile, conocida por su carácter iterativo e incremental, para avanzar sobre el análisis y diseño de forma adaptativa y ágil, permitiendo así una evolución continua y una entrega temprana de valor al cliente/usuario.

Líneas de Investigación y Desarrollo

El enfoque para resolver la problemática sobre el costo e inmediatez en la distribución de alimentos es colaborativo, en ese sentido es requerida la participación de todos los beneficiarios que cuenten con un móvil para el retiro y distribución de mercadería, lo que permitirá que entre ellos puedan asistirse. A partir de lo comentado anteriormente, se identificaron cinco líneas de investigación.

Distancia que el móvil estará dispuesto a alejarse de su domicilio (comedor/beneficiario):
La herramienta se diseñará para asegurar que los comedores beneficiarios que son asistidos por otros beneficiarios se encuentren dentro de un radio sugerido/acotado de distribución, garantizando así la posibilidad de entrega de mercadería por otro comedor beneficiario.

Capacidad del vehículo distribuidor:
La herramienta deberá tener en cuenta la capacidad del vehículo encargado de la distribución para garantizar que el volumen de mercadería transportado se ajuste a los límites operativos del mismo, asegurando la viabilidad del transporte.

Asociación entre el móvil (vehículo) y el comedor beneficiario:
La herramienta facilitará la asociación entre los vehículos de distribución y los comedores beneficiarios, esto permitirá que los comedores puedan llevar mercadería a otros comedores (dependiendo de la capacidad del vehículo). beneficiarios basándose en la priorización por necesidad, garantizando que los recursos lleguen a quienes más lo requieren.



Proceso de cálculo para priorización de beneficiarios (comedores).
Dado que los vehículos tienen un límite de capacidad, la herramienta a partir de indicadores deberá realizar el cálculo para priorizar a los beneficiarios (comedores) más necesitados. El objetivo es que a partir de una lista de posibles beneficiarios que se encuentran dentro de un rango de distancia, solo los mas necesitado puedan recibir la mercadería.

Asignación de Grupo de Afinidad GAF.
La herramienta será diseñada para permitir confeccionar grupos de afinidad, en este sentido los comedores podrán agruparse por los criterios identificados en el análisis.

Desarrollo e Innovación

El proyecto propone un **módulo de software** para optimizar la logística del Banco Alimentario de La Plata, reduciendo costos y mejorando la eficiencia en la distribución de alimentos. Se implementan **algoritmos de optimización** para calcular rutas eficientes considerando la capacidad de los vehículos y la proximidad de los comedores. Además, se utiliza un **modelo de priorización** automatizado que clasifica a

los beneficiarios según sus necesidades. La innovación principal radica en la creación del **Grupo de Afinidad (GAF)**, que fomenta la colaboración entre comedores con capacidad de transporte, optimizando recursos. La arquitectura modular y escalable permite futuras expansiones, mientras que la integración de **machine learning** mejora la precisión en la asignación de recursos.

Resultados Obtenidos/Esperados

La investigación se centra en el análisis, diseño y desarrollo de un módulo de software destinado a optimizar el costo de distribución de alimentos entre los comedores beneficiarios. Este módulo tiene como objetivo principal proporcionar una solución representativa y basada en datos empíricos, lo que permitirá una gestión innovadora y eficiente de la distribución de alimentos en el Banco Alimentario. Al emplear técnicas avanzadas de análisis de datos y modelado, se busca minimizar los costos operativos y mejorar la eficacia en la asignación de recursos, garantizando así un impacto positivo y sostenible para la Asociación Civil.

Una vez culminada la etapa de análisis y diseño del sistema de información para resolver la problemática de la distribución holística de alimentos dentro del Banco Alimentario, varios entregables son esenciales para implementar y evaluar el sistema. Estos entregables pueden incluir:

Documento de Requisitos del Sistema.
Un informe detallado que enumere todos los requisitos funcionales y no funcionales del sistema. Debe describir las necesidades específicas del Banco Alimentario y cómo el sistema abordará esas necesidades.

Diseño del Sistema.
Un documento que describe la arquitectura y el diseño del sistema. Debería incluir diagramas, modelos de datos, esquemas y especificaciones técnicas, destacando cómo se abordarán los requisitos identificados.

Prototipos o Maquetas.
Si es aplicable, prototipos o maquetas interactivas del sistema que permitan a los usuarios y al personal del Banco Alimentario visualizar cómo funcionará el sistema antes de su implementación.

Plan de Implementación.
Un plan que describe cómo se llevará a cabo la implementación del sistema, incluyendo una cronología, recursos necesarios y responsabilidades.

Plan de Pruebas.
Un plan que detalla cómo se probará el sistema para garantizar que funcione correctamente. Esto incluye pruebas de funcionalidad, rendimiento y seguridad.

Documentación del Usuario.
Manuales o guías para los usuarios finales que expliquen cómo utilizar el sistema de información.

Entrenamiento.
Un programa de capacitación para el personal del Banco Alimentario que les enseñe cómo usar el sistema de información de manera efectiva.

Base de Datos.
Un documento conteniendo el modelo de base de datos y una muestra de datos relevantes para el Banco Alimentario, como información sobre las restricciones y distribución de los alimentos.

Formación de Recursos Humanos

Durante el desarrollo de la investigación se busca:

1. Participación de estudiantes avanzados: Se contará con la colaboración de tres alumnos avanzados de la carrera de Licenciatura en Sistemas, quienes aportarán sus conocimientos y habilidades técnicas al proyecto.
2. Tesis de grado: Este proyecto servirá como base para la elaboración de tesis de grado por parte de los estudiantes avanzados, permitiéndoles profundizar en temas específicos relacionados con la investigación y el desarrollo del software.
3. Base de investigación futura: La ejecución de la herramienta generará una valiosa base de datos empíricos, la cual se explotará en investigaciones futuras. Esto permitirá la continuidad y ampliación del conocimiento en la optimización de la distribución de alimentos y en la aplicación de técnicas de análisis de datos en el ámbito de los sistemas.

Bibliografia

[1] Stuart, T. (2009). *Waste: Uncovering the global food scandal*. W. W. Norton & Company.

[2] Gordon, A. (2021). *Managing food safety and sustainability: Food preservation, process monitoring, and assurance*. Wiley-Blackwell.

[3] Pullman, M., & Wu, Z. (2012). *Food supply chain management: Economic, social, and environmental perspectives*. Routledge.

[4] Riches, G. (2018). *Food bank nations: Poverty, corporate charity and the right to food*. Routledge.

[5] Christopher, M. (2016). *Logistics and supply chain management* (5th ed.). Pearson.

[6] Vicari, S., & Emmanouilidis, C. (2020). *Lean Six Sigma in the food industry*. Springer.

[7] Drèze, J., & Sen, A. (1989). *Hunger and public action*. Oxford University Press.

[8] Pinstrup-Andersen, P., & Watson II, D. D. (2011). *Food policy for developing countries: The role of government in global, national, and local food systems*. Cornell University Press.

[9] Charlebois, S., Juhasz, M., Otten, S., & Van Westendorp, R. (2022). *Food loss and food waste: Causes and solutions*. Palgrave Macmillan.

[9] Siddiqui, M. W. (2019). *Managing food quality and safety: Global context*. Apple Academic Press.

Esquema de información para municipio con capacidad de Ciudad Inteligente Sostenible

Carlos Lugani, Mauro Cambarieri, Nicolas Garcia Martinez Juan Cruz Martines Luquez, Guillermo Malpeli

Universidad Nacional de Rio Negro. Sede Atlántica
Laboratorio de Informática Aplicada
{clugani, mcambarieri, ngarciamjcmluquez, gmalpeli}@unrn.edu.ar

RESUMEN

Este proyecto tiene como objetivo la definición y diseño de un esquema informático de relaciones para definir características de Ciudad Inteligente con la orientación de sostenibilidad en lo que respecta a: ámbito de emisiones de carbono, uso de recursos, crecimiento urbano, tratamiento de residuos, sistemas de agua y saneamiento, vías de comunicación y transporte.

En especial se desarrollará el esquema para municipios de ciudades pequeñas de 50.000 a 100.000 habitantes con características que les son propias. Se preve la implementación progresiva y teniendo en cuenta ajustes necesarios de acuerdo a las situaciones particulares de los mismos. El resultado esperado será un esquema de información apoyado por un sistema informático que facilitará los ingresos de la información y la generación de indicadores, contando con capacidad de compartir la misma por la comunidad. Se buscará obtener mejoras en la calidad de vida de los habitantes a través de un sistema de información que incluirá el monitoreo del nivel de eficiencia de la energía que es consumida y de los factores que reducirán la emisión de carbono. Se tiene en cuenta para el desarrollo del esquema el cumplimiento de los Objetivos de Desarrollo Sostenible (ODS) [1] y la generación de indicadores que sean comprendidos por los ciudadanos.

Palabras clave: Ciudades Inteligentes, , ODS, Sostenibilidad, Indicadores.

CONTEXTO

El cambio climático es ineludible y las acciones que tiendan a mejorar la situación actual deben ser fomentadas por toda la sociedad, promoviendo la educación de los ciudadanos y las iniciativas para proteger el medio ambiente donde se establecen. Se analizarán como guía las variables que se tienen en cuenta para el desarrollo de un modelo que dará lugar a un sistema de información con los siguientes Objetivos de Desarrollo Sostenible (ODS).

Al desarrollar soluciones que tienden a mejorar estos principios se afirma que la Sostenibilidad de las ciudades se verá beneficiada. Las acciones que se proponen deben analizarse dentro del concepto de Ciudades Inteligentes [2]. Una ciudad se considera inteligente cuando logra integrar capacidades en varios aspectos que están unidos por las tecnologías de información y las comunicaciones de forma tal que la calidad de vida de sus ciudadanos se ve beneficiada, se utilizan mejor sus recursos, se promueve el desarrollo sostenible y la gobernanza y se promueven la calidad de los servicios que gozan los ciudadanos. Pueden no cumplirse todos los aspectos mencionados pero la intención es que exista un plan de desarrollo para alcanzar metas establecidas.

La sostenibilidad está dada por las premisas:

- Gestión de residuos: promoviendo actividades y procesos de reciclado, reducción de plásticos y envases, reciclado y reutilización de materiales y promoción de las economías circulares.
- Gestión del agua: el uso eficiente del agua y los sistemas de distribución, promoviendo el uso responsable en edificios, espacios públicos y hogares, protegiendo las fuentes de agua potable, su explotación, y los ecosistemas acuáticos que se encuentren alrededor de las ciudades.
- Espacios verdes y biodiversidad: conocer las capacidades actuales y potenciales de

las áreas verdes y parques, así como los alrededores de la ciudad es importante para mejorar la calidad del aire y promover la salud de los habitantes, promoviendo la diversidad y la recuperación de especies nativas, creando corredores verdes e incrementando la cantidad de espacios verdes urbanos.

- Transporte sostenible: tanto el transporte público como privado debería ser sostenible, fomentando el uso de vehículos eléctricos, creando facilidades para la circulación de bicicletas y mejorando las capacidades del transporte público.
- Eficiencia energética: reducir el consumo de energía eléctrica en edificios y espacios públicos, privados y hogares, promoción de construcciones eficientes a través de menor utilización de energía y gastos de recursos para su mantenimiento.
- Planificación urbana sostenible: las ciudades se deben diseñar para que se promueva el uso del suelo urbano en forma inteligente, de forma tal que se permita la accesibilidad a los espacios verdes, a las área de trabajo y comercio, facilitando la conectividad y el transporte público, permitiendo el crecimiento planificado de la ciudad en el tiempo.

En este trabajo se desarrollarán capacidades en las ciudades en lo que respecta a la gestión de residuos y gestión del agua. De acuerdo a un informe del Banco Mundial los desechos a nivel mundial crecerán un 70 por ciento para 2050 3 lo cual es imposible de sostener con la forma de administrar los residuos actualmente en uso. Entendiendo que la Participación Ciudadana y la Educación serán dos factores fundamentales para el logro de las iniciativas planteadas.

A continuación, se presentan dos antecedentes que el grupo de investigación ha desarrollado en el último año:

- PI 40-C-1158 “Esquema de información para municipio con capacidad de Ciudad Inteligente Sostenible” desarrollado en el Laboratorio de Informática Aplicada, Sede Atlántica, Universidad Nacional del Río Negro (UNRN).
- Simposio de Informática en el Estado de las JAIIO 49 -2024 - “Esquema de una ciudad inteligente con orientación a la sostenibilidad desde una visión de municipio”

Los miembros del equipo de trabajo se han desarrollado en proyectos de investigación de la UNRN: “Conceptos y Modelos de Referencia para la Formulación de una Agenda de Gobierno Electrónico” y “Evaluación de Tecnologías de la Información y la Comunicación para el Desarrollo de Ciudades Inteligentes en Río Negro” que han permitido la transferencia de servicios digitales innovadores para diversos Organismos Públicos y Privados.

1. INTRODUCCIÓN

Se plantea el desarrollo de un esquema de gestione información relativa a las siguientes aceleradores de sostenibilidad:

- Economia circular y reciclaje a través de la gestión de la información: Para que se puedan conocer los actores de una economía circular deben conocerse las acciones, las necesidades de relación y los circuitos de los desechos o sobre existencias. La gestión de la información sobre todas las acciones requiere una plataforma abierta que permita a los interesados observar en forma transparente el flujo de objetos o desechos que se desarrolla en esa economía circular e intervenir en su tratamiento. Las empresas, organismos, oficinas deberían involucrarse en un proceso que identifique los ingresos y egresos de materiales que se utilizan y que son descartados en sus operaciones diarias. Esta fase es necesaria para identificar y cuantificar los tipos de residuos u objetos que son utilizados en forma normal y su posterior utilización, transformación y descarte. En esta etapa

tanto el municipio como las entidades que se denominen como reutilizadores se pondrán en contacto mediante la plataforma de información para continuar con el ciclo de vida de los materiales.

- Reducción de residuos: el rol de la educación es primordial para la reducción de residuos [3] y se relacionan con lo siguiente: Consumo responsable. Concientización en calificación de residuos. Formas de separación de residuos y de su disposición de acuerdo a los formatos de disposición final que se encuentren en la ciudad. Conocimiento de procesos de reciclado o reutilización En este sentido se considera que la actividad mas importante es la de educación, información de disponibilidad de facilidades, publicidad de actores en el proceso de reciclaje, definición de reglas municipales disponibles para los ciudadanos, intercomunicación de las personas para socializar sus soluciones. Para todas estas actividades se propone como principal y único punto de información el sistema propuesto ya que centralizará todas las acciones que tienden a mejorar la reducción de residuos además de poder cuantificar las acciones que se realicen ya que se encuentran registradas en el mismo sistema. La correcta disposición de residuos y su tratamiento correcto facilita el reciclaje, la existencia de la economía circular, reduce costos y por lo tanto afecta la economía y mejora las emisiones de carbono. [4]. [5]

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Esta investigación ayudará a la reducción de desechos, implicará una reducción y ahorro de superficies destinadas a la disposicion final de desperdicios, conduciendo a un ahorro en la gestión de residuos del municipio.

Se deben establecer los indicadores al comienzo del proyecto para considerar los ahorros en que se incurran. Asimismo se deben relacionar los indicadores entre si, por ejemplo, en caso que la reducción de desechos impacte en forma positiva sobre la calidad del agua y suelo [6]., y estos indicadores consecuentemente impacten en la calidad del aire. Estos análisis deben realizarse pero para esto se debe contar primero con un sistema que registre y ponga a disposición en forma abierta la informacion.

Objetivo Generales: El esquema propuesto tiene en cuenta el diseño de un portal y un sistema de información para presentar datos e información relacionada con:

- Indicadores para el municipio, teniendo en cuenta cantidad de habitantes y personas que ingresan en forma temporal a la región, así como factores como turismo.
- Acciones de gestión desarrolladas, las normales de recolección y reciclado y las especiales como reciclado de residuos específicos.
- Iniciativas municipales, de la industria, y de los ciudadanos
- Resultados esperados y obtenidos, en políticas, procesos e indicadores a desarrollar
- Indicadores de medio ambiente (la polución en el aire como un índice negativo y la forestación como un indice positivo relacionado con el indicador anterior).

Los indicadores finales son propios de la ciudad objetivo y sus características geográficas. La fuente de origen de los indicadores se visualiza como:

- Indicadores del Municipio
- Indicadores de dispositivos autónomos instalados en la región (dispositivos físicos que reciben y transfieren datos a través de redes inalámbricas IoT – Internet de las cosas o Internet of Thigs) [5] [7]
- Información de empresas y personas que ingresen en el portal con el objetivo de compartir situaciones, proyectos o iniciativas. [8]

El esquema se completa con una funcionalidad educativa y de concientización. Se proponen los siguientes aspectos:

- Educación inicial con participación en proyectos establecidos
- Educación secundaria con involucramiento en la creación de nuevos proyectos
- Concientización de toda la población
- Apoyo a iniciativas del sector privado, comercial, cooperativas y particulares

El sistema concentrará la información, sobre la base de un análisis previo interrelaciona los indicadores dando lugar a nuevos. De esta forma presentará un Monitor integrado con la característica de presentar también la cantidad de ciudadanos involucrados en proyectos, y las personas que hayan sido capacitadas en este sentido.

3. RESULTADOS OBTENIDOS/ ESPERADOS

Los resultados son esperados debido a que especialmente se diseñarán y desarrollará un esquema innovador para el tratamiento y concientización de información con las premisas que han sido mencionadas anteriormente. Asimismo se esperan en general los beneficios siguientes como resultados finales:

- Indicadores
- Sistema de información como monitor de indicadores
- Diseño de portal y acciones de socialibilización
- Material relacionado al manejo del sistema para diferentes niveles de educación

4. FORMACIÓN DE RECURSOS HUMANOS

El grupo de trabajo se encuentra formado por tres investigadores formados y dos investigadores en formación, todos docentes de la carrera Licenciatura en Sistemas. En el marco de este trabajo se desarrollará un plan de trabajo de la carrera de Doctorado y un plan de trabajo para una Maestría.

5. BIBLIOGRAFÍA

[1]. Programa de las Naciones Unidas para el Desarrollo (2024) https://www.undp.org/es/sustainable-development-goals

[2]. Estrategia argentina de Ciudades Inteligentes. Recuperado de: https://www.argentina.gob.ar/sites/default/files/estrategia_argentina_de_ciudades_inteligentes.pdf

[3]. https://www.bancomundial.org/es/news/press-release/2018/09/20/global-waste-to-grow-by-70-percent-by-2050-unless-urgent-action-is-taken-world-bank-report

[4]. E Behzad, B Wang, K Lewis, F Duarte, C Ratti, S Behdad, The future of waste management in smart and sustainable cities: A review and concept paper, Waste Management, Volume 81, 2018, Pages 177-195, ISSN 0956-053X, https://doi.org/10.1016/j.wasman.2018.09.047.

[5]. I. Sosunova and J. Porras, "IoT-Enabled Smart Waste Management Systems for Smart Cities: A Systematic Review," in *IEEE Access*, vol. 10, pp. 73326-73363, 2022, doi: 10.1109/ACCESS.2022.3188308.

[6]. Patil, R.A., Ramakrishna, S. A comprehensive analysis of e-waste legislation worldwide. Environ Sci Pollut Res 27, 14412–14431 (2020). https://doi.org/10.1007/s11356-020-07992-1

[7]. Nanda, S., Berruti, F. Municipal solid waste management and landfilling technologies: a review. Environ Chem Lett 19, 1433–1456 (2021). https://doi.org/10.1007/s10311-020-01100-y

[8]. Karak, T., Bhagat, R. M., & Bhattacharyya, P. (2012). Municipal Solid Waste Generation, Composition, and Management: The World Scenario. Critical Reviews in Environmental Science and Technology , 42 (15), 1509–1630. https://doi.org/10.1080/10643389.2011.569871

Estrategias para desacoplar funcionalidades y facilitar el desarrollo de Software en Instituciones Públicas: El caso de la Universidad Nacional de Río Negro

Patricio Nicolas Castro, Mauro Germán Cambarieri

Universidad Nacional de Rio Negro. Sede Atlántica
Laboratorio de Informática Aplicada
{pcastro, mcambarieri}@unrn.edu.ar

RESUMEN

Este proyecto tiene como propósito explorar y adoptar la arquitectura Vertical Slice Architecture (VSA), cuyo enfoque prioriza la organización del código en "slices" o segmentos verticales orientados a funcionalidades específicas. La VSA se presenta como una alternativa para instituciones públicas, al ofrecer un diseño centrado en el usuario y en los casos de uso en la que mejora significativamente la claridad, mantenibilidad y escalabilidad de los sistemas. A diferencia de las arquitecturas tradicionales basadas en capas horizontales, este modelo permite que cada funcionalidad se implemente de manera integral dentro de cada segmento, ya que abarca desde la interfaz de usuario hasta la capa de persistencia de datos. Este enfoque asegura una alineación directa con los procesos institucionales y la misión estratégica de la organización.

La VSA se caracteriza por su capacidad para encapsular todos los componentes necesarios (lógica de negocio, interfaces y acceso a datos) dentro de cada slice vertical. Esta modularidad facilita el aislamiento funcional, lo que permite realizar refactorizaciones, pruebas unitarias y de integración y reducir el riesgo de interferir con sistemas legados u otras funcionalidades del sistema. Tal cualidad resulta particularmente relevante en entornos públicos, donde la coexistencia con tecnologías heredadas y procesos burocráticos rígidos exige soluciones que minimicen el riesgo operativo durante actualizaciones o migraciones. Adicionalmente, la arquitectura admite un diseño escalable: funcionalidades simples se resuelven con estructuras simples, mientras que casos de uso complejos incorporan capas de abstracción progresivas, esto aumenta su flexibilidad y favorece la adaptabilidad ante cambios normativos o técnicos, recurrentes en el sector público.

Un contraste crítico con las arquitecturas multicapa tradicionales radica en la reducción de dependencias transversales. En modelos horizontales, componentes como bases de datos o servicios compartidos suelen generar acoplamientos que dificultan puntualmente las modificaciones. La VSA, en cambio, mitiga este riesgo al limitar el alcance de los cambios a slices específicos.

El trabajo también explora patrones arquitectónicos complementarios para optimizar la VSA, que incluyen Transaction Script, CQRS y Mediator. Estos modelos refuerzan la cohesión dentro de cada slice. También, se realiza una comparativa con la Arquitectura Hexagonal, se discuten ventajas, desventajas, inclusive la curva de aprendizaje en cada caso. Al finalizar se valida mediante un caso de estudio referido a un sistema de información de la Universidad Nacional de Río Negro con código de para ilustrar su implementación.

Palabras clave: Arquitectura de software, Transformación digital, Vertical Slice Architecture, Arquitectura Hexagonal, Mediator, CQRS.

CONTEXTO

En el contexto del Gobierno Digital, se emplean lenguajes de modelado específicos del dominio para representar el conocimiento del dominio y las reglas de negocio de las soluciones de software. Esto permite una mayor precisión y comprensión del dominio gubernamental y facilita el desarrollo de código (OMG, 2014) [1]. Esto brinda beneficios como la detección temprana de errores, la capacidad de realizar análisis y simulaciones y la mejora de la comunicación entre expertos en el dominio y desarrolladores de software.

La Arquitectura de software conforma la columna vertebral de cualquier sistema y constituye uno de sus principales atributos de calidad [2]. La Arquitectura de Software, es definida como “la

organización fundamental de un sistema, encarnada en sus componentes, las relaciones entre ellos y el ambiente y los principios que orientan su diseño y evolución" [3]. Distintos ejemplos como la Arquitectura Hexagonal - también conocida como Puertos y Adaptadores-, desarrollada por Alistair Cockburn [4], luego la Arquitectura de Datos, Contexto e Interacción, (DCI por sus siglas en inglés Data, Context and Interaction) de James Coplien y Trygve Reenskaug [5] son muy similares, con el mismo objetivo, centrarse en el dominio específico y, dividir el software en capas. Esto facilita la mantenibilidad e independencia de los frameworks, es decir, que la arquitectura no depende de la existencia de una librería o biblioteca de software. Testeable - en este caso, la Arquitectura puede probar las reglas del negocio sin la interfaz de usuario, base de datos, servidor web o, cualquier otro elemento externo. Independiente de la Interfaz de Usuario - el componente de la Arquitectura puede cambiar algo fácilmente, sin cambiar el resto del sistema. Una interfaz de usuario web podría ser reemplazada por una consola, por ejemplo, sin cambiar las reglas del negocio. Independiente de la base de datos - dado que las reglas de negocio no están ligadas a la base de datos, es posible cambiar el motor de base de datos. Independiente de cualquier agente externo - las reglas de negocio no conocen en absoluto de las interfaces con el mundo exterior.

En este escenario, los recursos tecnológicos se presentan tanto como facilitadores esenciales y herramientas de innovación para mejorar la calidad, el desarrollo de software en el ámbito público.

El presente trabajo se encuentra en el marco del Proyecto de Investigación 40-C-875 "Herramientas Informáticas de Dominio Específico para el Desarrollo de Servicios Digitales Innovadores para Comunidades Urbanas y Rurales en el Marco de Ciudades y Regiones Inteligentes" desarrollado en el Laboratorio de Informática Aplicada, Sede Atlántica, y la dirección de Sistemas de Rectorado de la Universidad Nacional del Río Negro (UNRN).

1. INTRODUCCIÓN

La Vertical Slice Architecture (VSA) es un enfoque de diseño que prioriza la organización del código en función de las características del negocio, en lugar de las capas técnicas tradicionales. Cada "slice" contiene todos los componentes necesarios para implementar una funcionalidad específica, desde la interfaz de usuario hasta el acceso a datos [6].

Esta arquitectura facilita la organización del proceso de desarrollo, ya que abarca tanto la codificación como las pruebas. La VSA implica implementar una porción de funcionalidad tras otra, de manera independiente y donde cada una abarca todas las capas del stack, desde la interfaz de usuario y la lógica de la aplicación hasta el almacenamiento de datos [7].

VSA se enfoca en los casos de uso y la interacción del usuario con el sistema [8]. Cada módulo implementa al menos un caso de uso, pero el enfoque no define detalladamente cómo realizar los cortes verticales, lo que permite adaptabilidad según las necesidades del proyecto y la experiencia del equipo Esto también permite elegir un diseño arquitectónico diferente por corte [8]. Como sugiere Bogard (2020) mientras que los casos de uso simples y directos podrían no requerir un diseño interno complejo, los casos de uso más grandes y complicados pueden estructurarse de una manera más completa. No solo permite la adaptación a las circunstancias actuales, sino también una refactorización más fácil después de, por ejemplo, añadir más casos de uso a un corte y cuando su diseño requiere más consideración.

A continuación se muestra un diagrama que muestra un corte vertical, Figura 1.

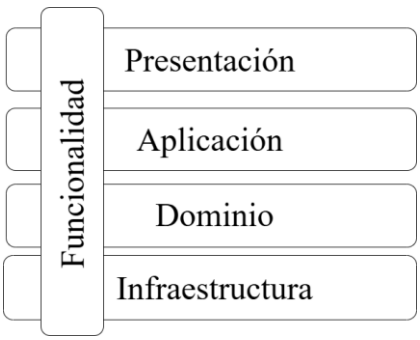


Figura 1. Vertical Slice Architecture. Adaptada de Vertical Slice Architecture [6].

Como ya se mencionó VSA propone dividir por funcionalidad, en lugar de por cuestiones técnicas. Uno de los beneficios de esta división es permitir que cada Request tenga un caso de uso distinto.

Esto nos lleva a aplicar el patrón CQRS [9], este es un patrón de diseño de software que separa la lógica de las aplicaciones entre las operaciones de escritura y las de lectura. O dicho de otra forma, las operaciones que producen cambios en el sistema y las operaciones que hacen consultas.

Otro patrón que se utiliza en esta arquitectura es Mediator, un patrón de diseño de comportamiento que se encuentra en el catálogo de GoF [10] y se encarga de controlar cómo interactúan los objetos entre sí. Aunque VSA no indique el uso de CQRS o Mediator, estos patrones fluyen bien juntos [11]. El objetivo de usar mediator es desacoplar las piezas entre sí. Mediator actúa como intermediario entre para la comunicación de los objetos se con la ayuda del mediator y ayuda a romper ayuda a romper el vínculo estrecho entre los colaboradores [11].

En el contexto de VSA, es utilizado Transaction Script, que organiza la lógica empresarial por procedimientos que maneja una única solicitud de la presentación. Esto implica obtener una entrada de datos, realizar algún procesamiento y una respuesta o persistir un cambio en el sistema. Por lo tanto, es un patrón simple que facilita el desarrollo inicial de una característica y permite que la arquitectura evolucione dada las necesidades del dominio.

El patrón Transaction Script organiza la lógica de un caso de uso como un único procedimiento que realiza llamadas directamente a la base de datos o a través de una capa delgada. Este patrón es básico y se utiliza cuando el comportamiento es simple, como obtener o guardar datos sin necesidad de un gran modelo de dominio [12]. Aplica cuando no necesitamos una gran complejidad y alcanza con asegurar que obtenemos los datos deseados para generar la respuesta.

Esto nos da la posibilidad de agregar complejidad solo cuando es necesario, cuando debemos establecer reglas y límites para proteger las invariantes y asegurar la consistencia de los cambios. Las validaciones de datos pueden ocurrir en el handler, y las validaciones que corresponden con reglas de negocio e invariantes tendrían que ocurrir en el handler. Lo que nos lleva a modelar el dominio, y representar reglas de comportamiento que se deben cumplir [13].

VSA se contrapone a arquitecturas como la arquitectura hexagonal, que busca independizar el núcleo de la aplicación de las dependencias externas mediante el uso de puertos y adaptadores. Mientras que la arquitectura hexagonal se enfoca en la independencia tecnológica y la testabilidad, VSA además busca la cohesión y la facilidad de desarrollo específico.

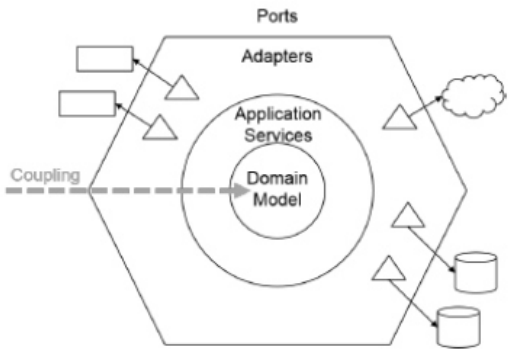


Fig. 2. Arquitectura Hexagonal de A. Cockburn [14]

La Arquitectura VSA se caracteriza por organizar el código por funcionalidad, donde cada slice contiene los componentes necesarios para implementar una funcionalidad específica. Esto resulta en un bajo acoplamiento entre slices y una alta cohesión dentro de cada uno. Por otro lado, la Arquitectura Hexagonal separa la aplicación entre la lógica de negocio y los puertos que definen las interacciones con el mundo exterior, lo que logra un fuerte desacoplamiento del núcleo de la aplicación respecto a tecnologías concretas.

En términos de complejidad, la VSA es generalmente más simple de entender e implementar, mientras que la Hexagonal puede ser más abstracta y compleja, requiere una comprensión clara del diseño del sistema. En cuanto a la abstracción, la VSA evita abstracciones prematuras, favorece la simplicidad y la evolución del código, mientras que la Hexagonal introduce abstracciones a través de puertos que definen interfaces para los adaptadores.

La flexibilidad en la VSA se manifiesta en el diseño de cada slice, adaptándose a las necesidades específicas de cada funcionalidad. En cambio, la Hexagonal ofrece alta flexibilidad para cambiar tecnologías externas, aunque puede ser menos adaptable a requisitos específicos de cada caso de uso.

En cuanto a la refactorización, la VSA requiere un esfuerzo continuo para evitar la duplicación de código, mientras que la Hexagonal facilita este proceso al permitir el intercambio de adaptadores sin afectar al núcleo. Finalmente, para los aspectos transversales como autenticación y logging, la VSA requiere consideraciones adicionales, mientras que la Hexagonal los maneja a través de adaptadores, lo que permite una implementación centralizada.

Finalmente, en VSA la curva de aprendizaje inicial puede ser más suave, ya que el enfoque en funcionalidades concretas facilita la comprensión del código, lo que requiere de equipos menos expertos. Sin embargo, a medida que el proyecto crece, es crucial tener un buen conocimiento de refactoring para evitar la duplicación de código y mantener la cohesión.

En cambio, en la arquitectura hexagonal, la curva de aprendizaje inicial puede ser más pronunciada, ya que requiere comprender los conceptos de puertos y adaptadores, patrones de diseño y la separación entre la aplicación y el mundo exterior. Sin embargo, una vez dominados estos conceptos, la arquitectura hexagonal proporciona una base sólida para construir aplicaciones testeables y mantenibles. Además, requiere de equipos relativamente expertos. A continuación, se mencionan las ventajas y desventajas de cada una.

En VSA se identifican las siguientes ventajas: (1) Facilidad de desarrollo y mantenimiento: Los cambios se limitan a un slice, reduce los side effects (secundarios o colaterales) y reduce el "context switching". (2) Testabilidad: Los slices aislados permiten pruebas unitarias y de integración sencillas. (3) Curva de aprendizaje más rápida: Los desarrolladores pueden ser productivos más rápidamente al centrarse en una sola característica.

En VSA se identifican las siguientes desventajas: (1) Duplicación de código: Si no se refactoriza adecuadamente, riesgo de código

duplicado en diferentes slices. (2) Complejidad general: A medida que crecen las features, puede volverse compleja si no se gestiona correctamente.

En la Arquitectura Hexagonal se identifican las siguientes ventajas: (1) Independencia tecnológica: Permite cambiar fácilmente las tecnologías externas (bases de datos, frameworks) sin afectar el núcleo de la aplicación. (2) Testabilidad: Facilita las pruebas unitarias al poder simular los adaptadores externos. (3) Mantenibilidad: Promueve un código limpio y modular, que facilita el mantenimiento a largo plazo.

En la Arquitectura Hexagonal se identifican las siguientes desventajas: (1) Complejidad inicial: Requiere una inversión inicial en diseño y comprensión de los conceptos de puertos y adaptadores. (2) Sobrediseño: Puede ser excesivamente abstracto si no se aplica con criterio.

El siguiente código de ejemplo está basado en un proyecto referido a un sistema de información de la UNRN. El mismo es un caso de uso para agregar un usuario, que muestra cómo implementar la arquitectura VSA con Spring Boot y Java, y aplica el patrón mediator.

```
@RestController
@RequestMapping("/command-users")
class UserAPICommand {

    JMediator mediator;

    public UserAPICommand(JMediator mediator) {
        this.mediator = mediator;
    }

    @PostMapping
    @ResponseStatus(CREATED)
    @Operation(summary = "Crear una cuenta", security
    void add(@RequestBody CreateUserCommand command)
        mediator.send(command);
    }

}
```

Fig. 3. Endpoint para agregar un usuario

```
public record CreateUserCommand(
        UUID id,
        String name,
        String lastname,
        String email,
        String username,
        String password

        ) implements IRequest<Void> {
}
```

Fig. 4. Objeto DTO del usuario

```
@Handler
public class CreateUserHandler {
    private final UsersResource usersResource;

    public CreateUserHandler(UsersResource usersResource) {
        this.usersResource = usersResource;
    }

    void create(CreateUserCommand command) {
        var representation = new UserRepresentation();
        representation.setEmail(command.email());
        representation.setUsername(command.email());
        representation.setFirstName(command.name());
        representation.setLastName(command.lastname());
        representation.setAttributes(Map.of("password", List.of(cc
        representation.setRequiredActions(List.of(UPDATE_PASSWORD.
        representation.setEnabled(true);
        try (var res = usersResource.create(representation)) {
            if (!Objects.equals(res.getStatus(), 200)) {
                throw new EmptyException("Error creating user");
            }
        } catch (Exception e) {
            throw new EmptyException("Error creating user");
        }
    }
}
```

Fig. 5. Lógica del caso de uso agregar usuario

```
@Validator
class createUserValidator {
    void validate(CreateUserCommand command) {
        if (!usersResource.searchByEmail(command.email(), true).isEmpty()) {
            throw new EmailAlreadyExistsException();
        }
        if (!usersResource.searchByUsername(command.username(), true).isEmpty()) {
            throw new UsernameAlreadyExistsException();
        }
    }
}
```

Fig. 6. Validaciones de datos

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El objetivo principal de la línea de investigación aquí presentada es:

Diseñar e implementar estrategias basadas en una arquitectura de vertical slice para facilitar el desarrollo de software de calidad en instituciones públicas, promover la modularidad, el desacoplamiento de funcionalidades y la escalabilidad de las soluciones tecnológicas.

Sus ejes específicos son: (a) Estudiar el estado del arte de enfoques modernos de desarrollo de software, que incluye la arquitectura de vertical slice, el testing automatizado (TDD - Desarrollo Guiado por Pruebas) y las prácticas de refactoring continuo, para identificar mejores prácticas aplicables al contexto de instituciones públicas, (b) Evaluar marcos de trabajo y herramientas que combinen la arquitectura de vertical slice con técnicas de testing automatizado y refactoring, con el fin mejorar la calidad del software y facilitar su mantenimiento y evolución en el contexto de instituciones públicas, (c) Diseñar estrategias de testing automatizado y prácticas de refactoring continuo dentro del ciclo de vida del desarrollo que se integren con la arquitectura de vertical slice, con el objetivo de mejorar la calidad del código, reducir

la deuda técnica y mantener la modularidad y el desacoplamiento de las funcionalidades, (d) Desarrollar un caso de estudio en el que se apliquen técnicas de testing automatizado (como TDD) y refactoring en un proyecto basado en vertical slice, con el fin de validar su efectividad en el contexto de instituciones públicas, se estudiará en el contexto de la Universidad Nacional de Río Negro, (e) Desarrollar un caso de estudio en un proyecto basado en vertical slice, que aplique técnicas de testing y refactoring, con el fin de validar su efectividad en el contexto de instituciones públicas, (f) Diseñar una arquitectura de referencia que permita su instanciación para la aplicación de los enfoques propuestos y promueva la reutilización sistemática de componentes y la evolución constante del software, (g) Desarrollar y validar prototipos de software, pasibles de ser implementados y transferidos al medio.

3. RESULTADOS OBTENIDOS/ ESPERADOS

Como resultado de este proyecto, se espera identificar arquitecturas de software, metodologías y modelos conceptuales para el desarrollo de software en dominios complejos, en busca de priorizar la modularidad y la evolución de los sistemas mediante el enfoque de Vertical Slice Architecture (VSA). Desde el punto de vista arquitectónico, el diseño basado en VSA posibilitará la creación de regiones o capas desacopladas, lo que favorecerá su evolución de manera independiente y sin estar restringidas por tecnologías específicas. Este modelo contribuirá significativamente a la modernización de sistemas legados y ofrecerá una alternativa flexible para la transformación digital de instituciones públicas.

Se espera que el software cumpla con lo siguiente:

- Alta flexibilidad y adaptación a cambios en reglas de negocio, ya que cada funcionalidad podrá evolucionar de manera aislada.
- Mayor mantenibilidad y modularidad, al estructurar el código en el que abarcan toda el "stack" tecnológico de cada funcionalidad,.
- Independencia tecnológica, permitir integrar diferentes tecnologías dentro de cada "slice".
- Capacidad de reutilización de componentes.
- Aumento en la calidad del software debido a la mayor facilidad en la prueba e implementación de nuevas funcionalidades
- Reducir la deuda técnica y mejorar la eficiencia en el desarrollo de software en el contexto de las instituciones públicas.
- Promover la confianza en las soluciones tecnológicas desarrolladas para instituciones públicas.

4. FORMACIÓN DE RECURSOS HUMANOS

El grupo de trabajo se encuentra integrado por un investigador formado, un investigador en formación y dos alumnos avanzados de la carrera Licenciatura en Sistemas. En su marco se desarrollarán prácticas profesionales supervisadas, becas de formación práctica y se producirán dos trabajos finales de carrera de grado y la elaboración de trabajo de tesis de la carrera de Maestría en Ingeniería de Software de la UNLP.

5. BIBLIOGRAFÍA

[1] OMG. (2014). OMG Systems Modeling Language (OMG SysML™) Version 1.4. Object Management Group

[2] Clements, P., et al, "Software Architecture in Practice", Pearson Education, (2003).

[3] IEEE Standards Association, "1471-2000 - IEEE Rec ommended Practice for Architectural Description for Software-Intensive Systems", available at: http://standards.ieee.org/findstds/standard/1471-2000.html

[4] Alistair Cockburn. Alistair.Cockburn.us, Hexagonal architecture. https://alistair.cockburn.us/hexagonal-architecture/ (2023)

[5] James O Coplien, Trygve Mikkjel Heyerdahl Reenskaug: "The data, context and interaction paradigm. Proceedings of the 3rd annual conference on Systems, programming, and applications: software for humanity." October 2012 Pages 227–228 https://doi.org/10.1145/2384716.2384782

[6] Bogard, J. (2018). Vertical slice architecture. Jimmy Bogard. http://jimmybogard.com/vertical-slice-architecture

[7] Esposito, D. (2024). Clean architecture with .NET. Addison Wesley.

[8] J. Bogard, NDC Porto Software Developers Conference. "Vertical Slice Architecture." Porto, 24.04.2020. [Online]. Available: Vertical Slice Architecture - Jimmy Bogard https://www.youtube.com/watch?v=5kOzZz2vj2o

[9] M. Fowler. CQRS. http://martinfowler.com/bliki/CQRS.html, 2011.

[10] E. Gamma, R. Helm, R. Johnson, and J. Vlissides, Design Patterns: Elements of Reusable Object-Oriented Software, Reading, MA: Addison-Wesley, 1995.

[11] Marcotte, C. H., & Zebdi, A. (2022). An Atypical ASP. NET Core 6 Design Patterns Guide: A SOLID adventure into architectural principles and design patterns using. NET 6 and C# 10. Packt Publishing Ltd.

[12] M. Fowler. Transaction Script. http://martinfowler.com/eaaCatalog/transactionScript.html, 2003

[13] Cambarieri, M., Difabio, F., & García Martínez, N. (2020). Implementación de una Arquitectura de Software guiada por el Dominio. In XXI Simposio Argentino de Ingeniería de Software (ASSE 2020)-JAIIO 49 (Modalidad virtual).

[14]. A. Cockburn, "The Pattern: Ports and Adapters," 2005: disponible en: https://alistair.cockburn.us/hexagonal-architecture/ [accedido: 18/08/2020]

El valor del uso de los datos aplicado a la estrategia alimentaria

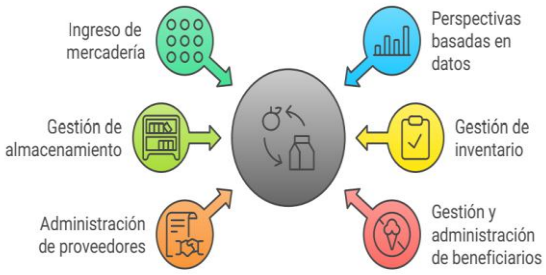
Ing. Franco Quintero [1,7] C.C Marcelo Taruschio [2,7]
Dr. Edgar Gonzalo Cossio Franco[3,8] AS David Esteban Clai[4,7], AS Analía Forte[5,7,] AS Nicolas Campetti[6,7]

[7]Facultad de Ciencias Exactas e Ingeniería. Universidad Católica de La Plata.
[8]Universidad del Valle de Atemajac, UNIVA, México
[1] francopablo.quintero@ucalpvirtual.edu.ar [2] marcelo.taruschio@ucalp.edu.ar
[3]edgar.cossio@univa.mx [4] foxblue70@gmail.com [5] analia_fff@hotmail.com
[6] nicolas_campetti@hotmail.com

Resumen

En el marco del proyecto "Estrategia de distribución holística de alimentos como herramienta de gestión. Su aplicación en el Banco Alimentario", esta investigación tiene como objetivo delinear el estado de situación y la problemática actual del Banco Alimentario de la Ciudad de La Plata. En ese sentido se identifican varias áreas de estudio cruciales, tales como, gestión de proveedores, ingreso y egreso de mercadería, gestión de inventarios y la distribución de mercadería para que, a partir del análisis y diseño se pueda crear una herramienta de software (modulo) que permita mejorar el resultado de distribución de alimentos a los beneficiarios (comedores) del Banco Alimentario, el objetivo es reducir el costo de distribución de alimentos.

Se presentan avances sobre la investigación realizada. El propósito general, en una primera etapa es identificar y especificar los requerimientos para el desarrollo de una herramienta de software que permita reducir o mitigar los costos asociados a la distribución de alimentos y mercaderías a los diferentes beneficiarios (comedores), identificando y priorizando aquellos con mayores necesidades.



De manera adicional se generará como producto entregable un documento detallado que describe la estrategia de distribución holística, incluyendo:

- Visión general de la estrategia.
- Análisis de la situación actual.
- Objetivos y principios de la estrategia.
- Estrategias específicas para optimizar la logística, la gestión de inventario, utilizando herramientas de IA.
- Plan de acción detallado.
- Indicadores clave de rendimiento (KPIs).

Palabras clave:
Gestión de proyectos de software. Logística. IA aplicado a software de logística. Análisis de datos aplicado a la estrategia alimentaria.

Contexto

La investigación surge como una línea de trabajo colaborativa con la organización Civil Banco Alimentario de la ciudad de La Plata. Ante el creciente avance de la necesidad de satisfacer la distribución y recolección de los alimentos con la mayor inmediatez, esta investigación se enfoca en enfrentar los desafíos asociados a los costos de adquisición y distribución.

Si bien el resultado del trabajo está enfocado en la distribución de los alimentos, el análisis y diseño se realizó sobre todo el estado de situación (íntegro) del Banco Alimentario.

A continuación, la metodología y pasos aplicados:

• Recopilación de información sobre las mejores prácticas en distribución holística de alimentos, análisis del contexto del Banco Alimentario y sus necesidades específicas.

• Evaluación de la capacidad actual de distribución, identificación de brechas y oportunidades de mejora.

• Desarrollo de un plan de acción detallado que incluya estrategias para optimizar la logística, la gestión de inventario, utilizando herramientas de IA y explotación de datos obtenidos del relevamiento.

• Puesta en marcha del plan de acción, seguimiento y evaluación de resultados.



Nota: El comedor beneficiario es el único responsable de buscar la mercadería asignada por el Banco Alimentario.

Introducción

En el contexto global actual, la inseguridad alimentaria persiste como un problema grave, mientras grandes cantidades de alimentos comestibles se desperdician anualmente. Organizaciones como los bancos de alimentos han surgido para enfrentar este desafío, aunque a menudo lo hacen sin un enfoque estratégico y holístico. Este proyecto propone optimizar la gestión de recursos, reducir el desperdicio de alimentos y mejorar la ayuda a la comunidad, potenciando la sostenibilidad a largo plazo de la organización desde la óptica de la distribución de alimentos.

Siguiendo la naturaleza del desarrollo de un producto de software, la estrategia se forja a partir de la identificación de un mapa de impacto, que sienta las bases de las necesidades de los requerimientos y expectativas de los usuarios. A partir de las necesidades se definirá un roadmap de historias de usuarios hacia el logro del producto deseado. Con las historias de usuarios identificadas se creará el “release plan” junto con su MVP (Mínimo Producto Viable) como punto de partida y se empleará una metodología Agile, conocida por su carácter iterativo e incremental, para avanzar sobre el análisis y diseño de forma adaptativa y ágil, permitiendo así una evolución continua y una entrega temprana de valor al cliente/usuario.

Líneas de Investigación y Desarrollo

El enfoque para resolver la problemática sobre el costo e inmediatez en la distribución de alimentos es colaborativo, en ese sentido es requerida la participación de todos los beneficiarios que cuenten con un móvil para el retiro y distribución de mercadería, lo que permitirá que entre ellos puedan asistirse. A partir de lo comentado anteriormente, se identificaron cinco líneas de investigación.

Distancia que el móvil estará dispuesto a alejarse de su domicilio (comedor/beneficiario):
La herramienta se diseñará para asegurar que los comedores beneficiarios que son asistidos por otros beneficiarios se encuentren dentro de un radio sugerido/acotado de distribución, garantizando así la posibilidad de entrega de mercadería por otro comedor beneficiario.

Capacidad del vehículo distribuidor:
La herramienta deberá tener en cuenta la capacidad del vehículo encargado de la distribución para garantizar que el volumen de mercadería transportado se ajuste a los límites operativos del mismo, asegurando la viabilidad del transporte.

Asociación entre el móvil (vehículo) y el comedor beneficiario:
La herramienta facilitará la asociación entre los vehículos de distribución y los comedores beneficiarios, esto permitirá que los comedores puedan llevar mercadería a otros comedores (dependiendo de la capacidad del vehículo). beneficiarios basándose en la priorización por necesidad, garantizando que los recursos lleguen a quienes más lo requieren.



Proceso de cálculo para priorización de beneficiarios (comedores).
Dado que los vehículos tienen un límite de capacidad, la herramienta a partir de indicadores deberá realizar el cálculo para priorizar a los beneficiarios (comedores) más necesitados. El objetivo es que a partir de una lista de posibles beneficiarios que se encuentran dentro de un rango de distancia, solo los mas necesitado puedan recibir la mercadería.

Asignación de Grupo de Afinidad GAF.
La herramienta será diseñada para permitir confeccionar grupos de afinidad, en este sentido los comedores podrán agruparse por los criterios identificados en el análisis.

Desarrollo e Innovación

El proyecto propone un **módulo de software** para optimizar la logística del Banco Alimentario de La Plata, reduciendo costos y mejorando la eficiencia en la distribución de alimentos. Se implementan **algoritmos de optimización** para calcular rutas eficientes considerando la capacidad de los vehículos y la proximidad de los comedores. Además, se utiliza un **modelo de priorización** automatizado que clasifica a

los beneficiarios según sus necesidades. La innovación principal radica en la creación del **Grupo de Afinidad (GAF)**, que fomenta la colaboración entre comedores con capacidad de transporte, optimizando recursos. La arquitectura modular y escalable permite futuras expansiones, mientras que la integración de **machine learning** mejora la precisión en la asignación de recursos.

Resultados Obtenidos/Esperados

La investigación se centra en el análisis, diseño y desarrollo de un módulo de software destinado a optimizar el costo de distribución de alimentos entre los comedores beneficiarios. Este módulo tiene como objetivo principal proporcionar una solución representativa y basada en datos empíricos, lo que permitirá una gestión innovadora y eficiente de la distribución de alimentos en el Banco Alimentario. Al emplear técnicas avanzadas de análisis de datos y modelado, se busca minimizar los costos operativos y mejorar la eficacia en la asignación de recursos, garantizando así un impacto positivo y sostenible para la Asociación Civil.

Una vez culminada la etapa de análisis y diseño del sistema de información para resolver la problemática de la distribución holística de alimentos dentro del Banco Alimentario, varios entregables son esenciales para implementar y evaluar el sistema. Estos entregables pueden incluir:

Documento de Requisitos del Sistema.
Un informe detallado que enumere todos los requisitos funcionales y no funcionales del sistema. Debe describir las necesidades específicas del Banco Alimentario y cómo el sistema abordará esas necesidades.

Diseño del Sistema.
Un documento que describe la arquitectura y el diseño del sistema. Debería incluir diagramas, modelos de datos, esquemas y especificaciones técnicas, destacando cómo se abordarán los requisitos identificados.

Prototipos o Maquetas.
Si es aplicable, prototipos o maquetas interactivas del sistema que permitan a los usuarios y al personal del Banco Alimentario visualizar cómo funcionará el sistema antes de su implementación.

Plan de Implementación.
Un plan que describe cómo se llevará a cabo la implementación del sistema, incluyendo una cronología, recursos necesarios y responsabilidades.

Plan de Pruebas.
Un plan que detalla cómo se probará el sistema para garantizar que funcione correctamente. Esto incluye pruebas de funcionalidad, rendimiento y seguridad.

Documentación del Usuario.
Manuales o guías para los usuarios finales que expliquen cómo utilizar el sistema de información.

Entrenamiento.
Un programa de capacitación para el personal del Banco Alimentario que les enseñe cómo usar el sistema de información de manera efectiva.

Base de Datos.
Un documento conteniendo el modelo de base de datos y una muestra de datos relevantes para el Banco Alimentario, como información sobre las restricciones y distribución de los alimentos.

Formación de Recursos Humanos

Durante el desarrollo de la investigación se busca:

1. Participación de estudiantes avanzados: Se contará con la colaboración de tres alumnos avanzados de la carrera de Licenciatura en Sistemas, quienes aportarán sus conocimientos y habilidades técnicas al proyecto.
2. Tesis de grado: Este proyecto servirá como base para la elaboración de tesis de grado por parte de los estudiantes avanzados, permitiéndoles profundizar en temas específicos relacionados con la investigación y el desarrollo del software.
3. Base de investigación futura: La ejecución de la herramienta generará una valiosa base de datos empíricos, la cual se explotará en investigaciones futuras. Esto permitirá la continuidad y ampliación del conocimiento en la optimización de la distribución de alimentos y en la aplicación de técnicas de análisis de datos en el ámbito de los sistemas.

Bibliografia

[1] Stuart, T. (2009). *Waste: Uncovering the global food scandal*. W. W. Norton & Company.

[2] Gordon, A. (2021). *Managing food safety and sustainability: Food preservation, process monitoring, and assurance*. Wiley-Blackwell.

[3] Pullman, M., & Wu, Z. (2012). *Food supply chain management: Economic, social, and environmental perspectives*. Routledge.

[4] Riches, G. (2018). *Food bank nations: Poverty, corporate charity and the right to food*. Routledge.

[5] Christopher, M. (2016). *Logistics and supply chain management* (5th ed.). Pearson.

[6] Vicari, S., & Emmanouilidis, C. (2020). *Lean Six Sigma in the food industry*. Springer.

[7] Drèze, J., & Sen, A. (1989). *Hunger and public action*. Oxford University Press.

[8] Pinstrup-Andersen, P., & Watson II, D. D. (2011). *Food policy for developing countries: The role of government in global, national, and local food systems*. Cornell University Press.

[9] Charlebois, S., Juhasz, M., Otten, S., & Van Westendorp, R. (2022). *Food loss and food waste: Causes and solutions*. Palgrave Macmillan.

[9] Siddiqui, M. W. (2019). *Managing food quality and safety: Global context*. Apple Academic Press.

Fortaleciendo la Conexión entre Proyectos de Software y el Bienestar Comunitario

Mg. Rodolfo Bertone [1] C.C Marcelo Taruschio [2]
Lic. Martin Correa[3] AS Federico Agustin Maza[4], AS Perrino Julian[5,] AS Forciniti Federico[6,] AS Giordano María Agostina[7]

Facultad de Ciencias Exactas e Ingeniería. Universidad Católica de La Plata.
[1] rodolfo.bertone@ucalp.edu.ar [2] marcelo.taruschio@ucalp.edu.ar
[3]martin.correa@ucalp.edu.ar [4]federico.maza@ucalpvirtual.edu.ar [5]jperrino182@gmail.com
[6] fforciniti921@gmail.com [7]agostinagiordano@hotmail.com.ar

Resumen

El concepto de "Usina de Ideas" se ha consolidado como un espacio, tanto virtual como físico, que impulsa la generación, evaluación y desarrollo de ideas dentro de la comunidad universitaria. A través del grupo de investigación y transferencia de la Facultad de Ciencias Exactas e Ingeniería, este entorno se ha convertido en un motor para la innovación y el trabajo colaborativo dentro de la institución.

A lo largo de 2024, el proyecto ha avanzado en la consolidación de sus líneas de investigación y transferencia, poniendo el foco en la aplicación del conocimiento y su impacto en la sociedad. En este contexto, se han llevado adelante diversas investigaciones cuyos avances han sido presentados en congresos especializados y continúan en desarrollo.

Actualmente, el trabajo se orienta a diversas iniciativas. Se está desarrollando un prototipo de sistema para el seguimiento de legajos digitales basado en tecnología Blockchain, lo que permitirá mejorar la trazabilidad y seguridad en la gestión documental. También se está explorando una solución digital para optimizar el uso compartido de espacios de trabajo, fomentando nuevas dinámicas de colaboración. En el ámbito del Internet de las Cosas (IoT), se avanza en la implementación de un sistema de control de temperatura para la producción de cerveza artesanal, con el objetivo de optimizar la eficiencia y calidad del proceso. Además, se está trabajando en un sistema inteligente de asistencia para la compra y venta de inmuebles, diseñado para mejorar la toma de decisiones mediante herramientas avanzadas de análisis de datos. Cada una de estas iniciativas refuerza el compromiso del proyecto con la innovación y la transferencia tecnológica, contribuyendo tanto al ámbito académico como al sector productivo.

Palabras clave:
Gestión de proyectos de software. Automatización y Control. Internet de las cosas. Gobernanza electrónica

Contexto

Esta línea de investigación se basa en avances previos, consolidando resultados y ampliando su aplicabilidad. Desde su inicio, ha sido un esfuerzo colaborativo entre docentes y alumnos, fortaleciendo tanto la investigación como las competencias de los estudiantes. Su formalización responde a un plan estratégico que promueve la producción

científica y la transferencia de conocimiento, asegurando su continuidad con recursos específicos.

Además de la investigación, este espacio permite a los estudiantes aplicar sus conocimientos en experiencias finales antes de graduarse, participando en proyectos que generan aplicaciones concretas y publicables. En un contexto tecnológico desafiante, estos proyectos fomentan la investigación aplicada y la transferencia tecnológica, integrando equipos multidisciplinarios para desarrollar soluciones innovadoras y efectivas. [1][2][3]

Introducción

En el ámbito académico, la investigación y la transferencia de conocimiento juegan un rol básico, no solo en el desarrollo de soluciones innovadoras, sino también en la consolidación de la relación entre la universidad y su entorno social y productivo. La capacidad de una institución educativa privada y católica para generar conocimientos aplicados que resuelvan problemas reales es un aspecto clave en la formación de los futuros profesionales y en el impacto de la UCALP en el entorno social en el que se encuentra inmersa. En este contexto, nuestra institución ha diseñado la "Usina de Ideas", un espacio académico y de colaboración que facilita la producción, evaluación y difusión de conocimientos, promoviendo una interacción constante entre alumnos, docentes y diversos actores de la comunidad, todos con el objetivo de aportar soluciones tecnológicas y sociales a los desafíos actuales. [4][5][6]

La "Usina de Ideas" se enfoca en una serie de iniciativas que buscan no solo mejorar el conocimiento académico, sino también generar soluciones tecnológicas con impacto práctico en la sociedad. En este contexto, actualmente se están desarrollando proyectos de gran relevancia, como un sistema para el seguimiento de legajos digitales basado en tecnología Blockchain, que permitirá optimizar la trazabilidad y la seguridad en la gestión de documentos. Además, se exploran soluciones digitales para mejorar la gestión de espacios de trabajo compartidos, fomentando nuevas dinámicas colaborativas. En el ámbito del Internet de las Cosas (IoT), se avanza en la implementación de un sistema de control de temperatura para la producción de cerveza artesanal, con el objetivo de mejorar la eficiencia y la calidad del proceso productivo. Por otro lado, se está trabajando en un sistema inteligente para la compra y venta de inmuebles, que utilizará herramientas avanzadas de análisis de datos para mejorar la toma de decisiones. Estos proyectos son ejemplos de cómo los alumnos, al participar en estos desarrollos. A través de estas actividades, la "Usina de Ideas" proporciona un entorno que refuerza tanto la formación académica como la transferencia de conocimiento, involucrando a alumnos, docentes y otros actores de la comunidad en la construcción de un futuro más tecnológico y colaborativo.

Líneas de Investigación. Objetivos.

Este proyecto, que continúa el trabajo iniciado el año anterior, tiene como objetivo principal la creación de un entorno que favorezca la definición y desarrollo de subproyectos con aplicaciones concretas y transferibles a la sociedad. Los subproyectos se centran en el desarrollo de soluciones basadas en

software, aunque, en etapas más avanzadas, se contemplan también el uso de tecnologías de hardware, como Arduino o Raspberry, para aplicaciones específicas.

Para alcanzar estos objetivos, en primer lugar, se busca facilitar la interacción entre alumnos y docentes a través de reuniones de brainstorming, con el propósito de fomentar la generación de ideas innovadoras que puedan resolver problemas específicos. En segundo lugar, se promoverá la elaboración de casos de estudio detallados para evaluar la viabilidad y el impacto potencial de los subproyectos, considerando tanto los aspectos técnicos como los económicos y, eventualmente, sociales. Estos casos servirán como base para presentar formalmente los diferentes subproyectos.

El proceso de desarrollo de los subproyectos incluirá la documentación de los avances, desafíos y resultados obtenidos, lo que permitirá crear informes técnicos, monografías y materiales que podrán ser presentados en congresos y eventos académicos. Cada subproyecto contará con un director encargado de supervisar su desarrollo, quien pondrá a consideración de los participantes los avances y resultados obtenidos, asegurando la participación activa de todos los involucrados.

Actualmente, las iniciativas en las que se está trabajando están dirigidas a la creación de soluciones tecnológicas que tengan una aplicación en distintos sectores. Como se dijo anteriormente, un ejemplo de esto es el desarrollo de un sistema de gestión de legajos digitales utilizando Blockchain, con el objetivo de mejorar la seguridad y trazabilidad en el manejo de documentos. Además, se está investigando una plataforma digital que permita gestionar de manera más eficiente el uso compartido de espacios de trabajo, lo cual promueve nuevas formas de colaboración entre equipos. En cuanto al Internet de las Cosas (IoT), se está desarrollando un sistema de control de temperatura para la producción de cerveza artesanal, con el fin de optimizar los procesos y garantizar la calidad del producto final. Por otro lado, se está creando un sistema inteligente para asistir en la compra y venta de propiedades, utilizando herramientas avanzadas de análisis de datos para facilitar la toma de decisiones en este ámbito.[7][8]

Estas iniciativas buscan no solo generar soluciones innovadoras, sino también contribuir a la transferencia de conocimiento en áreas clave, como la digitalización de procesos en el ámbito gubernamental y el fortalecimiento de organizaciones no gubernamentales (ONGs). De este modo, se trabaja en proyectos que tienen un impacto directo en la sociedad, mejorando diversos sectores y promoviendo el desarrollo sostenible.[9]

La participación en estos sectores representa una oportunidad única para canalizar los resultados de la investigación hacia problemas concretos y de impacto real en la comunidad. La "Usina de Ideas" no solo proporciona un espacio para la innovación tecnológica, sino que también favorece la colaboración multidisciplinaria, creando un puente entre la academia y las necesidades de la sociedad. De este modo, el proyecto no solo tiene un enfoque académico, sino que también refuerza su relevancia social y contribuye activamente al desarrollo sostenible y al bienestar colectivo, consolidando su impacto en el ámbito académico y más allá.

Desarrollo e Innovación

Los proyectos en curso dentro del marco de la "Usina de Ideas" se centran en la investigación y desarrollo de soluciones innovadoras y transferibles a la comunidad. Estos proyectos abordan problemas específicos en diversas áreas como la gestión documental, la colaboración en espacios de trabajo, la optimización de procesos productivos y la toma de decisiones en el mercado inmobiliario, incorporando tecnologías emergentes como Blockchain, el Internet de las Cosas (IoT) y la inteligencia de datos.

Sistema de Seguimiento de Legajos Digitales Basado en Blockchain

El desarrollo de un sistema de seguimiento de legajos digitales tiene como objetivo optimizar la gestión documental mediante el uso de las tecnologías Blockchain. Este enfoque se fundamenta en la capacidad de Blockchain para garantizar la trazabilidad, seguridad e inmutabilidad de los datos, características esenciales para la gestión de documentos sensibles en entornos digitales. Las características principales que debe tener este sistema son: Seguridad y Autenticidad, Interoperabilidad, Escalabilidad y Adaptabilidad, Accesibilidad y Usabilidad, Cumplimiento Normativo.

Solución Digital para el Uso Compartido de Espacios de Trabajo

La optimización de los espacios de trabajo es un desafío clave en el contexto actual, donde la colaboración y la flexibilidad en el uso de los espacios son esenciales. Este proyecto busca desarrollar una solución digital que permita gestionar y optimizar el uso de estos espacios, promoviendo nuevas dinámicas de colaboración en ambientes laborales. Las características que debe incorporar esta solución son: Gestión Eficiente de Recursos, Integración con Herramientas de Gestión, Interfaz Amigable, Análisis de Datos para Mejora Continua, Escalabilidad y Flexibilidad.

Sistema de Control de Temperatura para la Producción de Cerveza Artesanal (IoT)

La implementación de un sistema de control de temperatura en la producción de cerveza artesanal es un ejemplo de aplicación del Internet de las Cosas (IoT) para optimizar procesos productivos. Este sistema debe permitir la supervisión y control de las condiciones ambientales durante el proceso de producción, mejorando la calidad del producto final. Las características esenciales de este sistema incluyen: Monitoreo en Tiempo Real, Automatización y Control Remoto, Integración con sensores IoT, Escalabilidad y Adaptabilidad, Análisis Predictivo y Optimización.

Sistema Inteligente de Asistencia para la Compra y Venta de Inmuebles

El sistema inteligente de asistencia para la compra y venta de inmuebles busca facilitar la toma de decisiones en el sector inmobiliario mediante el análisis avanzado de datos. Las características de este sistema incluyen: Recopilación y Análisis de Datos, Toma de Decisiones Basada en Inteligencia Artificial, Interfaz de Usuario Intuitiva, Simulaciones y Pronósticos, Accesibilidad y Transparencia.

Resultados y Objetivos

El impacto de esta experiencia se refleja en logros concretos, por ejemplo, se publicó en 2024 el trabajo denominado "Traductor de lenguaje de señas simplificado por una inteligencia artificial." [10]. Se están preparando diversas publicaciones científicas para distintos congresos, donde se compartirán avances en Blockchain para gestión documental, IoT en procesos productivos y otras áreas clave. Más allá de la difusión científica, esto fortalece el posicionamiento del proyecto y abre nuevas oportunidades de colaboración e innovación.

Formación de Recursos Humanos

El proyecto no solo impulsa la investigación, sino que también tiene como objetivo formar nuevos Recursos Humanos entre los alumnos de la carrera que participan. Cinco estudiantes completaron su formación y obtuvieron su título de grado gracias a su trabajo en estas iniciativas, aplicando sus conocimientos en desafíos reales. Además, un docente está finalizando su tesis de Maestría, consolidando el vínculo entre la producción científica y la capacitación avanzada. Más que un espacio de investigación, la *Usina de Ideas* se ha convertido en una plataforma de crecimiento profesional y desarrollo tecnológico con impacto real.

Referencias

[1] Freeman, C. (1995). "The 'National System of Innovation' in historical perspective." *Cambridge Journal of Economics, 19*(1), 5-24.

[2] Von Hippel, E. (2005). *Democratizing Innovation*. MIT Press.

[3] Chesbrough, H. (2003). *Open Innovation: The New Imperative for Creating and Profiting from Technology*. Harvard Business Press.

[4] Etzkowitz, H., & Leydesdorff, L. (2000). "The dynamics of innovation: From National Systems and 'Mode 2' to a Triple Helix of university–industry–government relations." *Research Policy, 29*(2), 109-123.

[5] Perkmann, M., & Walsh, K. (2007). "University–industry relationships and open innovation: Towards a research agenda."*International Journal of Management Reviews, 9*(4), 259-280.

[6] Cohen, W. M., Nelson, R. R., & Walsh, J. P. (2002). "Links and impacts: The influence of public research on industrial R&D." *Management Science, 48*(1), 1-23.

[7] Tapscott, D., & Tapscott, A. (2016). *Blockchain Revolution: How the Technology Behind Bitcoin and Other Cryptocurrencies is Changing the World*. Penguin.

[8] Singh, M., & Duhan, P. (2020). *Internet of Things (IoT): A comprehensive review of the technology and its applications*. *Journal of Ambient Intelligence and Humanized Computing, 11*(9), 3857-3874. https://doi.org/10.1007/s12652-020-01791-0

[9] United Nations. (2020). *The UN E-Government Survey 2020: Digital Government in the Decade of Action for Sustainable Development*. United Nations. Recuperado de https://www.un.org/development/desa/dpad/publication/un-e-government-survey-2020/

[10] 15327 Traductor de lenguaje de señas simplificado por una inteligencia artificial. https://cacic2024.info.unlp.edu.ar/workshops/

GAMECORE: UN FRAMEWORK BACKEND PARA VIDEOJUEGOS AUTORITATIVOS

Emilio Ormeño[1], Cintia Ferrarini Oliver[2]
1. Instituto de Informática, 2. Departamento de Informática,
Facultad de Ciencias Exactas, Físicas y Naturales, UNSJ
eormeno@gmail.com, ferrarinicintia@gmail.com

RESUMEN

GameCore es un framework backend desarrollado para ofrecer una plataforma robusta y escalable destinada a la implementación de videojuegos autoritativos. Su diseño integra mecanismos avanzados de sincronización, autenticación, gestión de estados y eventos, orientados a garantizar la integridad y seguridad en entornos multijugador. La implementación con Laravel y Eloquent, y la adopción de patrones de diseño modernos (Componentes, State, inyección de servicios y callbacks de ciclo de vida), posiciona a GameCore como una solución integral para desarrolladores y estudios que deseen optimizar la experiencia del usuario.

CONTEXTO

El trabajo forma parte del proyecto CICITCA denominado "Framework USIM para el desarrollo de Aplicaciones Web", aprobado según resolución 2902-CS-UNSJ. La iniciativa surge ante la creciente demanda de infraestructuras que aseguren el control autoritativo en videojuegos multijugador, permitiendo gestionar de forma centralizada la simulación del juego y garantizando una experiencia segura y homogénea [1].

1. INTRODUCCIÓN

Los videojuegos multijugador requieren infraestructuras de backend que aseguren la coherencia, baja latencia y seguridad en la transmisión de datos [2]. Los sistemas autoritativos son esenciales para prevenir fraudes y discrepancias en la simulación. GameCore se presenta como un framework innovador que integra estos requerimientos, facilitando a los desarrolladores la implementación de sistemas complejos sin sacrificar escalabilidad ni extensibilidad [3], [4].

1.1. LA PLATAFORMA GAMECORE

GameCore es una solución integral que permite la implementación de videojuegos autoritativos a través de un conjunto de herramientas orientadas a la autenticación, gestión centralizada del estado, sincronización en tiempo real y manejo de eventos [3]. La plataforma se integra de forma nativa con diversos motores de videojuegos, permitiendo a los desarrolladores concentrarse en la lógica de juego mientras el framework gestiona la robustez y seguridad del entorno multijugador [5].

1.2. ARQUITECTURA DE GAMECORE

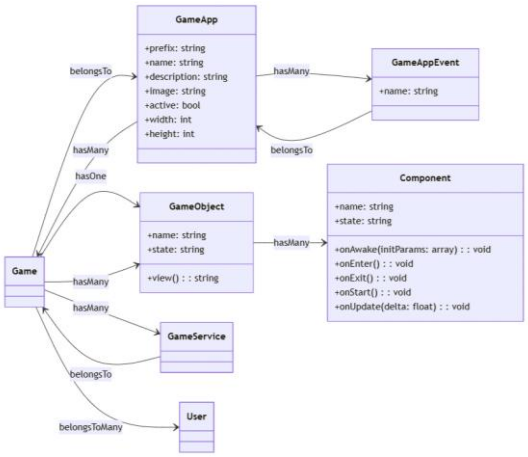
El framework se implementa siguiendo una arquitectura cliente-servidor, en la que el servidor actúa como la autoridad central de la simulación. Entre sus características distintivas se encuentran:

- *Gestión de Microservicios y Persistencia*: GameCore utiliza Laravel y Eloquent para gestionar microservicios y garantizar la persistencia de todos los elementos del videojuego, asegurando la integridad y consistencia de la información [6].
- *Patrón de Componentes*: Inspirado en la arquitectura de Unity3D [7], el framework implementa el patrón de Componentes [8], permitiendo que un Ga-

meObject posea uno o más componentes que definen su funcionalidad. Cada instancia de GameObject [9] y sus componentes se persiste mediante Eloquent [10], lo que facilita implementaciones extensibles y escalables [5].

- *Patrón State para Máquinas de Estados*: Una característica distintiva es la capacidad de convertir los GameObjects en manejadores de contextos de una máquina de estados, asociando "estados" a los componentes [11]. Esto posibilita la implementación de objetos con estados complejos (por ejemplo, un personaje que puede estar caminando, saltando, corriendo, muriendo, etc.) y garantiza que dichos estados se gestionen de forma persistente [12].
- *Servicios de Juego Persistentes*: El framework permite definir Servicios de Juego inyectados de forma transparente y persistente. Estos servicios facilitan la implementación limpia de funcionalidades comunes en todo el juego, promoviendo la reutilización de código y la modularidad [13].
- *Sistema Persistente de Eventos y Suscripción Automática*: GameCore incorpora un sistema de eventos persistente, con suscripción automática que posibilita el desacoplamiento de funcionalidades. Los componentes pueden subscribirse a eventos, permitiendo una comunicación interna eficiente y flexible [1].
- *Ciclos de Vida de Componentes*: Los componentes de GameCore transitan por distintos ciclos de vida. Los desarrolladores pueden implementar los comportamientos en métodos callback específicos.
- *Prefabs para GameObjects*: Siguiendo la analogía con Unity3D, los GameObjects pueden empaquetarse en Prefabs, que actúan como plantillas estáticas con definiciones completas de estructuras (padres, hijos y componentes), facilitando la reutilización y la organización de escenas complejas [5].
- *Definición de GameApps*: El framework permite la creación de una o más aplicaciones de juego (GameApps), que definen los atributos generales de un videojuego, tales como:
    - Número máximo y mínimo de jugadores (por ejemplo, un juego single player: máximo = 1 y mínimo = 1).
    - Dimensiones del panel de juego (width y height).
    - Un prefab raíz encargado de generar los elementos esenciales del juego.
    - Conjunto de recursos, componentes y servicios específicos.

Además, GameCore gestiona tanto partidas individuales para cada usuario como partidas compartidas entre múltiples jugadores [14]. La siguiente figura muestra la interrelación entre los elementos más importantes del dominio.

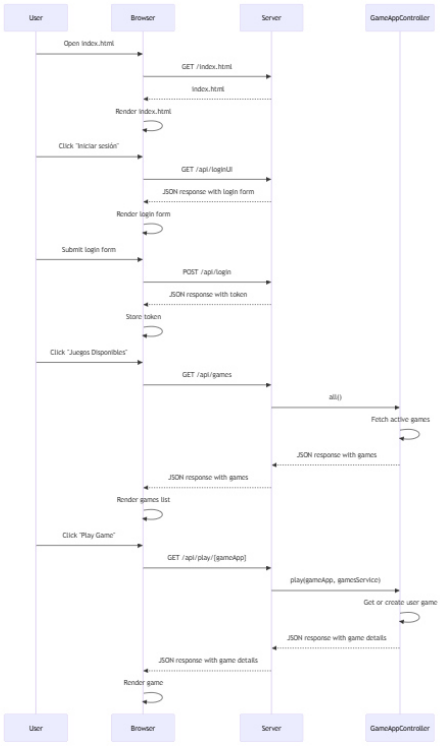


1.3. SERVICIO DE SINCRONIZACIÓN

El servicio de sincronización de GameCore

coordina las interacciones entre clientes y servidores, validando las acciones de los jugadores y actualizando el estado del juego en tiempo real. Esta arquitectura bidireccional minimiza la latencia y asegura que cada cambio se refleje de manera inmediata y precisa en todas las instancias conectadas [2].

La siguiente figura muestra cómo los distintos elementos interactúan entre sí a lo largo de una conexión entre un navegador y el servidor Laravel.



1.4. MANEJO DE EVENTOS Y ENTIDADES

GameCore incorpora un sistema modular para la gestión de eventos y entidades, facilitando la administración de colisiones, interacciones y la lógica de negocio mediante estructuras de datos optimizadas. La suscripción automática a eventos permite desacoplar funcionalidades, ofreciendo a los desarrolladores una mayor flexibilidad para implementar la dinámica del juego [1].

1.5. ESCENARIOS Y DESAFÍOS DE ESCALABILIDAD

En GameCore la arquitectura del framework está diseñada para enfrentar los desafíos inherentes a la escalabilidad en entornos multijugador. La implementación de balanceadores de carga y estrategias de replicación de datos, garantiza la continuidad operativa incluso en condiciones de alta concurrencia, manteniendo una experiencia de juego fluida y segura [14].

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

La línea de investigación y desarrollo de GameCore se centra en la optimización del rendimiento y la seguridad en videojuegos autoritativos. Entre las iniciativas prioritarias se destacan:

- Integración de microservicios y arquitectura basada en contenedores [4].
- Optimización de algoritmos de sincronización para la reducción de latencia [3], [2].
- Desarrollo de protocolos de seguridad avanzados para la protección de datos [1].
- Incorporación de inteligencia artificial para el monitoreo y detección temprana de anomalías [5].
- Ampliación de la funcionalidad de servicios y eventos para facilitar la modularidad y el desacoplamiento [13].

Estas líneas de trabajo se orientan a posicionar a GameCore como una solución innovadora en la industria del gaming, garantizando tanto la robustez del sistema como la experiencia de usuario.

3. RESULTADOS

Los ensayos piloto en producción[1] de GameCore han demostrado resultados significativos en términos de reducción de latencia y mejora en la estabilidad del estado del juego. Entre los beneficios observados se destacan:

- Según el videojuego piloto, se observan latencias de menos de 100 milisegundos. Y menos de 50 milisegundos en actualizaciones de GameObjects.
- Una gestión eficiente y centralizada de la persistencia de datos y estados [6].
- Flexibilidad y escalabilidad en la implementación de GameObjects y sus componentes [4], [5].
- Capacidad para gestionar servicios de juego y eventos de forma modular y desacoplada [13], [1].
- Soporte para la creación de configuraciones personalizadas (GameApps) que se adaptan a diversas modalidades de juego, desde partidas individuales hasta entornos multijugador compartidos [14].

Estos resultados brindan indicios sobre la viabilidad de GameCore como herramienta para el desarrollo de videojuegos autoritativos, potenciando tanto la experiencia de juego como la productividad en el proceso de desarrollo.

4. FORMACIÓN DE RECURSOS HUMANOS

El desarrollo e implementación de GameCore ha fortalecido la formación de recursos humanos especializados en tecnologías de videojuegos y desarrollo de software autoritativo. En la actualidad, se está dirigiendo un trabajo final de grado en la carrera de Licenciatura en Ciencias de la Computación, a cargo del alumno Martín Gómez, y una tesis de maestría en Informática, del Licenciado Alberto Domínguez.

Asimismo, se han llevado a cabo talleres de capacitación en formato de game jams, orientados a la puesta en práctica del framework. Estas iniciativas formativas han sido respaldadas por el Ministerio de Producción de la Provincia de San Juan, consolidando así una comunidad de aprendizaje y desarrollo que impulsa la innovación y el perfeccionamiento en el ámbito de los videojuegos.

5. REFERENCIAS

[1] S. Fernández and M. Ortega, “Seguridad y Autenticación en Videojuegos Online”,” *IEEE Security & Privacy*, vol. 18, no. 5, pp. 42–50, 2023.

[2] D. Kim Y and E. Zhao, “Optimización de la Latencia en Sistemas Multijugador”,” *Journal Real-Time Systems*, vol. 9, no. 3, pp. 115–128, 2020.

[3] “Sincronización en Tiempo Real en Juegos Multijugador”,” *IEEE Trans. Network Gaming*, vol. 12, pp. 234–249, 2022.

[4] R. Lin, “Microservicios y Escalabilidad en Plataformas de Juegos”,” in *en ACM SIGCOMM*, 2022, pp. 67–74.

[5] “Integración de Backend con Motores de Videojuegos”,” *GameDev Magazine*, vol. 11, pp. 34–39, 2022.

[6] “Laravel - the PHP framework for web artisans.” Available: https://laravel.com. [Accessed: Feb. 23, 2025]

[7] W. Li, “Animaton: Scriptable finite automaton for animation design in Unity3D game engine,” in *2022 4th International Conference on Pattern Recognition and Intelligent Systems*, New York, NY, USA: ACM, Jul. 2022. doi: 10.1145/3549179.3549196. Available: http://dx.doi.org/10.1145/3549179.3549196

[8] E. Aly, S. Elsawah, H. H. Turan, and M. J.

[1] https://damogame.com con el framework Laravel optimizado según guidelines y sugerencias de expertos.

Ryan, “Designing a pattern language to enhance model composability and reusability: An example with component-based probabilistic models,” *Environ. Model. Softw.*, vol. 169, no. 105807, p. 105807, Nov. 2023, doi: 10.1016/j.envsoft.2023.105807. Available: http://dx.doi.org/10.1016/j.envsoft.2023.105807

[9] S. B. Uzayr, “Working with scenes and GameObjects,” in *Mastering Unity*, Boca Raton: CRC Press, 2022, pp. 73–144. doi: 10.1201/9781003214755-3. Available: http://dx.doi.org/10.1201/9781003214755-3

[10] S. Sinha, “Database Migration and Eloquent,” in *Beginning Laravel*, Berkeley, CA: Apress, 2019, pp. 113–166. doi: 10.1007/978-1-4842-4991-8_5. Available: http://dx.doi.org/10.1007/978-1-4842-4991-8_5

[11] B. Bafandeh Mayvan, A. Rasoolzadegan, and Z. Ghavidel Yazdi, “The state of the art on design patterns: A systematic mapping of the literature,” *J. Syst. Softw.*, vol. 125, pp. 93–118, Mar. 2017, doi: 10.1016/j.jss.2016.11.030. Available: http://dx.doi.org/10.1016/j.jss.2016.11.030

[12] M. Pérez, “Modelos de Gestión de Estados en Videojuegos Autoritativos”,” *Comput. Graphics Forum*, vol. 31, no. 1, pp. 89–103, 2021.

[13] H. Ali, G. Mussbacher, and J. Kienzle, “Perspectives to promote modularity, reusability, and consistency in multi-language systems,” *Innov. Syst. Softw. Eng.*, vol. 18, no. 1, pp. 5–37, Mar. 2022, doi: 10.1007/s11334-021-00425-3. Available: http://dx.doi.org/10.1007/s11334-021-00425-3

[14] K. Roberts, “Estrategias de Replicación y Balanceo de Carga en Videojuegos”,” *IEEE Trans. Parallel Distributed Systems*, vol. 29, no. 7, pp. 210–225, 2023.

Implementaciones Tecnológicas en Ciudades Inteligentes

Rocío A. Rodríguez[1,2], Pablo M. Vera[1,2], Nicolas Repetti[1], Dora Mendoza[1,3], Claudia G. Alderete[1,2]

[1]Centro de Altos Estudios en Tecnología Informática (CAETI)
Facultad de Tecnología Informática
Universidad Abierta Interamericana (UAI)
Montes de Oca 745, Ciudad Autónoma de Buenos Aires, Argentina

[2]Facultad Regional Haedo (FRH)
Universidad Tecnológica Nacional (UTN)
Paris 522, Haedo, Buenos Aires, Argentina

[3]Laboratorio de Investigación y Desarrollo Tecnológico en Energías Renovables
Sede Regional Orán
Universidad Nacional de Salta, Argentina

{rocioandrea.rodriguez, pablomartin.vera, claudia.alderete}@uai.edu.ar
dmendoza412@oran.unsa.edu.ar
nicolas.repetti@alumnos.uai.edu.ar

RESUMEN

El crecimiento poblacional en las ciudades trajo consigo nuevos retos entre ellos analizar las posibilidades desde la infraestructura actual de las ciudades, la gestión de la movilidad de los ciudadanos, el impacto al medio ambiente. Estos son sólo ejemplos de áreas en las que es de importancia considerar a las TIC (Tecnologías de la Información y Comunicación) en beneficio de los ciudadanos.

En este artículo se presentan algunas implementaciones propuestas, vinculadas con distintas dimensiones de las ciudades inteligentes.

Palabras clave: Ciudades Inteligentes, Medio Ambiente, Seguridad Ciudadana, TIC

CONTEXTO

Esta línea de investigación y desarrollo (I+D) forma parte de los proyectos radicados en el Centro de Altos Estudios en Tecnología Informática (CAETI) de la Universidad Abierta Interamericana (UAI). En este proyecto participan docentes y alumnos, tanto de sede Centro como de la Castelar (ambas en la provincia de Buenos Aires). El proyecto cuenta con financiamiento asignado y una duración de 2 años.

1. INTRODUCCIÓN

"En un mundo cada vez más interconectado y tecnológicamente avanzado, el concepto de ciudades inteligentes se presenta como un punto de referencia clave para el cambio y la transformación en la manera en que concebimos y experimentamos la vida urbana" [1].

La sobrepoblación en áreas urbanas hace que se requiera implementar soluciones tecnológicas que faciliten las actividades cotidianas de sus habitantes, "243 millones de estadounidenses se concentran en el 3% urbano de ese país. En Tokio y sus alrededores, el área metropolitana más productiva del mundo, viven 36 millones de personas. En el centro de Bombay residen 12 millones de personas [...] En un planeta que dispone de enormes cantidades de espacio, preferimos las ciudades. Pese a que en la

actualidad sea más barato recorrer grandes distancias, cada vez más gente vive en las grandes áreas urbanas" [2].

"Las ciudades inteligentes se definen como aquellas que utilizan tecnologías de la información y la comunicación (TIC) para mejorar la calidad de vida de sus ciudadanos, gestionar los recursos de manera eficiente y fomentar el desarrollo sostenible. Este enfoque implica una visión integrada y multidisciplinaria, en la que se consideran aspectos económicos, sociales y ambientales para garantizar un crecimiento urbano equilibrado y sostenible" [3].

"En primer lugar, es fundamental reconocer el papel central que desempeñan las ciudades como motores del conocimiento en la era contemporánea" [1].

El término “Smart Cities” es actualmente muy buscado en Google. Puede analizarse con Google Trends que el interés a lo largo del tiempo en los últimos 5 años se mantiene (ver figura 1). En la figura 1, la herramienta muestra con una barra vertical que en un momento se modificó la forma en que recolectan los datos para poder mostrar las tendencias en las búsquedas de los usuarios.

Ordenados por tendencia del término “Smart Cities” (porcentaje de búsquedas del término frente al volumen de búsquedas de cada país), en los 15 primeros lugares aparecen los siguientes países: (1) Brasil; (2) Alemania; (3) Francia; (4) Italia; (5) Rusia; (6) Oman; (7) Puerto Rico; (8) Estados Unidos; (9) Corea del Sur; (10) Reino Unido; (11) Uzbekistan; (12) España; (13) Chile; (14) **Argentina;** (15) República Dominicana.

En Argentina la Secretaría de Modernización de la Nación (Presidencia de la Nación Argentina) ha establecido un modelo de ciudad inteligente [4], [5]. Este modelo surge del análisis de otros modelos previos existentes: ISO 37.120 [6]; ONU ITU [7]; European SCI [8]; ONU Habitat [9], [10].

El modelo establece 5 dimensiones y ejes de acción sobre cada una de ellas:

(1) Gobernanza:
- (a) Gobierno Abierto;
- (b) Infraestructura y Capacidades;
- (c) Plataforma de Servicios.

(2) Ambiente:
- (a) Calidad Ambiental;
- (b) Ecología Urbana y Resiliencia
- (c) Gestión de Recursos.

(3) Desarrollo Humano:
- (a) Educación;
- (b) Salud;
- (c) Seguridad;
- (d) Sociedad.

(4) Planeamiento Urbano:
- (a) Espacio Público;
- (b) Movilidad;
- (c) Transporte.

(5) Competitividad:
- (a) Contexto;
- (b) Innovación;
- (c) Productividad.

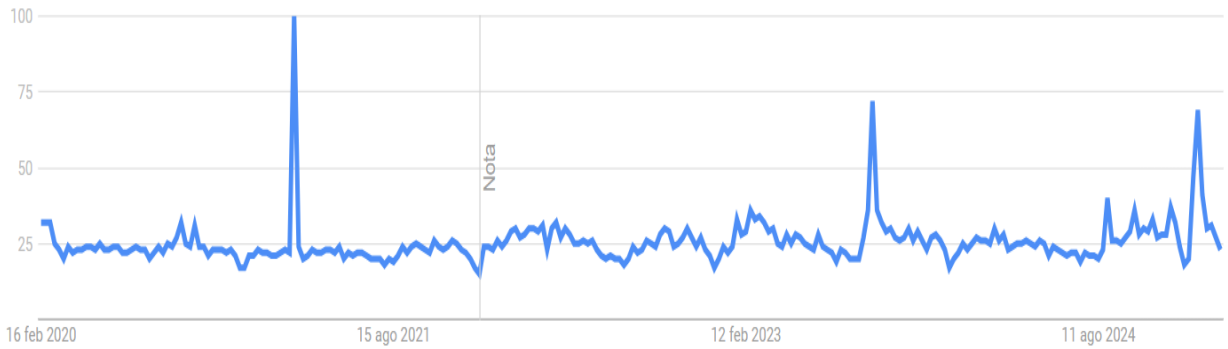


Figura 1. Popularidad del término Smat Cities en los últimos 5 años, según el volumen de búsquedas.

Es importante considerar las posibilidades de implementación de soluciones tecnológicas, en las distintas dimensiones, que permitan mejorar la calidad de vida de los ciudadanos.
En este proyecto se ha considerado principalmente 3 dimensiones: Ambiente, Planeamiento Urbano y Desarrollo Humano.

2. LÍNEAS DE INVESTIGACIÓN y DESARROLLO

Los ejes principales del trabajo son:

- Analizar las implementaciones existentes para Smart Cities (SC) y clasificarlas por dimensiones
- Comparar los planes de acción de los gobiernos nacionales de América Latina con respecto a las SC
- Comparar los puntos considerados por los distintos rankings de Ciudades Inteligentes
- Analizar formas de evaluar el grado en que las ciudades han implementado características de las SC
- Desarrollar estrategias e implementar recursos innovadores que permitan mejorar distintos aspectos propios de las SC

3. RESULTADOS OBTENIDOS/ESPERADOS

Durante el año pasado se ha trabajado mayormente en la dimensión de Desarrollo Humano (eje: Seguridad) y producto de ello surgieron dos publicaciones:

- Se propuso una implementación de bajo costo para el reconocimiento de patentes, solución realizada en conjunto con un grupo de Seguridad Vial, “Reconocimiento de Patentes: Implementación de Gendarmería Nacional en el Grupo Seguridad Vial Uruguai (Misiones)”, artículo presentado en CACIC (Congreso Argentino de Ciencias de la Computación) [11].
- Se presentó un artículo sobre el uso de datos abiertos para la toma de decisiones en JAIIO (Jornadas Argentinas de Informática) dicho artículo fue seleccionado entre los mejores “Seguridad Ciudadana en Ciudades Inteligentes: Uso de Datos Abiertos para la Toma de Decisiones” [12].

Se encuentran dos artículos en proceso de revisión en revistas indexadas uno de ellos como extensión del artículo presentado en JAIIO y el otro de ellos vinculado con la dimensión Ambiental, el cual se realizó en conjunto con la colaboración de docentes investigadores del Laboratorio de Investigación y Desarrollo Tecnológico en Energías Renovables (Sede Regional Orán, Universidad Nacional de Salta, Argentina).

También durante el año pasado fueron presentadas y defendidas satisfactoriamente dos tesis de grado de la carrera: Licenciatura en Gestión de Tecnología Informática de UAI (Universidad Abierta Interamericana), vinculadas con el presente proyecto:

- Análisis de la Aplicación Móvil "BA 147" en el Contexto de Ciudades Inteligentes – Tesista: Ezequiel Calelo [12].
- RFID Aplicado a la Seguridad Vial para la Prevención de Accidentes – Tesista: Sebastián Manai[13].

Y se encuentra en proceso de evaluación una tesis presentada para la misma carrera de grado: “Estrategias para mejorar la posición de Buenos Aires en el ranking de ciudades inteligentes: Un Enfoque Comparativo” (tesista: Federico Azansa).

Como puede observarse es una temática que puede abordarse desde diversos enfoques en cada uno de sus ejes de acción.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo está formado por 5 docentes todos con estudios de posgrado realizados ó en curso: 2 de ellos son doctores en Ciencias Informática graduados en la Universidad Nacional de La Plata (UNLP), 2 con maestría y actualmente realizando el Doctorado en Informática en la Universidad Abierta Interamericana (UAI) y 1 realizando la Maestría en Gestión Educativa (UAI).

Este proyecto también cuenta con la participación de alumnos de grado y posgrado de la UAI que no están vinculados con actividades docentes (actualmente en el proyecto colaboran 3 alumnos).
En esta área se encuentran en realización 1 tesis de doctorado (UNLP) y 1 tesina de grado (UAI), siendo directores de dichas tesis miembros del equipo de investigación.

5. BIBLIOGRAFIA

[1] Osorio, E. D. F. (2024). El camino hacia ciudades inteligentes: evolución tecnológica y urbana. Ciencia Vital, 2(2).

[2] Caragliu, A., Del Bo, C. and Nijkamp, P. (2009) Smart Cities in Europe.

[3] Alcalá Casillas, M. G., & Granados Macías, C. A. (2024). Ciudades inteligentes en la era del conocimiento. Derecho global. Estudios sobre derecho y justicia, 9(26), 133-157.

[4] Castella L. (2017). "La Importancia de un Modelo de Planificación Estratégica para el Desarrollo de Ciudades Inteligentes". Secretaría de Modernización, Presidencia de la Nación. JATIC 2017

[5] Palacio Cortés C. (2019). "Gobernanza: base en la Estrategia de Ciudades Inteligentes. El caso Argentino de País Digital". Dirección de Investigación, Innovación y Control Subsecretaría de Modernización de la Nación, Presidencia de la Nación. Smart City Expo Curitiba 2019

[6] ISO (2014), Desarrollo de Comunidades Sostenibles y Resilientes: Indicadores Globales para los servicios de las ciudades y la calidad de vida.

[7] ITU-T (2015), Grupo de Estudio 5, Informe Técnico sobre Ciudades Inteligentes y Sostenibles.

[8] Giffinger, Rudolf et al. (2007-2015), European Smart Cities, Universidad Tecnológica de Viena

[9] ONU Hábitat III (2015), United Nations Conference on Housing and Sustainable Urban Development, ONU.

[10] ONU Hábitat III, Objetivo 11 (2015) United Nations Conference on Housing and Sustainable Urban Development, ONU.

[11] Pesado, P. M., Thomas, P. J., & Chichizola, F. (2024). Libro de Actas-30° Congreso Argentino de Ciencias de la Computación-CACIC 2024. "Reconocimiento de Patentes: Implementación de Gendarmería Nacional en el Grupo Seguridad Vial Uruguai (Misiones)". Página 1321

[12] Repetti, N., & Rodriguez, R. (2024). Seguridad Ciudadana en Ciudades Inteligentes: uso de datos abiertos para la toma de decisiones. JAIIO, Jornadas Argentinas de Informática, 10(13), 34-45.

[13] Manai, S. E. (2024). RFID aplicado a la seguridad vial para la prevención de accidentes.

Plataforma inteligente de recomendaciones para optimizar la experiencia del usuario

Susana Chavez[1], Adriana E. Martín[1], Sergio Flores[1], Sara Zogbe[1]

[1]Departamento e Instituto de Informática - F.C.E.F. y N. - U.N.S.J.
Complejo Islas Malvinas. Cereceto y Meglioli. 5400. Rivadavia. San Juan, 0264 4234129

schavez@iinfo.unsj.edu.ar, arianamartinsj@gmail.com, sergior@gmail.com, sarazogbe@yahoo.com.ar, adrianamartin1@gmail.com

Resumen

En la era digital, la inteligencia artificial (IA) y las SuperApps han transformado la manera en que los usuarios interactúan con los servicios digitales, redefiniendo la personalización, accesibilidad y eficiencia en múltiples sectores. La IA ha permitido la automatización de tareas, el análisis predictivo de datos y la optimización de la experiencia del usuario mediante asistentes virtuales, chatbots y sistemas de recomendación avanzados.

Por otro lado, las SuperApps han evolucionado como ecosistemas digitales que integran múltiples servicios en una sola plataforma, facilitando desde pagos y comercio electrónico hasta transporte y asistencia personalizada. Empresas líderes como WeChat, Grab y Rappi han demostrado el potencial de este modelo, eliminando la necesidad de múltiples aplicaciones independientes y ofreciendo una experiencia más fluida e interconectada.

La combinación de IA y SuperApps ha dado lugar a aplicaciones más inteligentes y adaptativas, capaces de comprender y anticipar las necesidades de los usuarios. La próxima generación de aplicaciones móviles integrará estrategias avanzadas de seguridad, tecnologías inmersivas, SuperApps e IA.

La presente propuesta de investigación tiene como objetivo explorar y analizar los beneficios de integrar una SuperApp con tecnologías de la IA como machine learning, geolocalización y computación serverless para ofrecer recomendaciones dinámicas y adaptativas a las preferencias del usuario.

Palabras clave: Mobile Computing, IA, SuperApps, Computación Inmersiva, backend.

Contexto

El presente trabajo se encuadra dentro del área de I/D Innovación en Sistemas de Software y es una de las líneas de investigación internas, del proyecto: Soporte Serverless para aplicaciones móviles de nueva generación, cuya propuesta está en aprobada para el período 2023-2025. Así mismo el grupo de investigadores viene trabajando en proyectos relacionados con la computación móvil y distribuida desde hace más de 22 años. Como continuación del proyecto anterior: Computación Serverless para tratamiento de datos provenientes de dispositivos de IoT, se continúa el trabajo con investigadores de otras universidades, lo cual favorece notablemente a todos las instituciones participantes.

Introducción

La IA y la computación en la nube han revolucionado múltiples sectores, y el turismo no es la excepción. Sin embargo, aún existen pocas soluciones integrales que combinen IA generativa, geolocalización, asistentes

virtuales y sistemas de recomendación en una sola aplicación.

Con la expansión del acceso a dispositivos móviles y redes 5G, el uso de SuperApps basadas en arquitectura serverless permite ofrecer servicios escalables y accesibles globalmente, sin requerir dispositivos de alto rendimiento.

La implementación de IA en turismo puede impulsar la competitividad de destinos, mejorar la gestión del flujo de visitantes y potenciar la economía local al dirigir recomendaciones personalizadas a negocios y servicios turísticos.

En resumen, abordar este problema con una SuperApp basada en IA no solo mejora la experiencia del usuario, sino que también representa una oportunidad de innovación tecnológica en un sector con gran impacto económico y social.

Itinerario

Durante el periodo 2023 se comenzó a trabajar en el diseño y desarrollo de un prototipo de SuperApp turística que permite la incorporación de contenido multimedia audiovisual llamada “Itinerario”. Dicha aplicación incorpora un mapa interactivo con geolocalización del recorrido habitual que se realiza asociado a la actividad turística. Desde allí se accede a otras apps como la red Tulum (aplicación de transporte de colectivo de la provincia de San Juan) para tener una alternativa de llegar al lugar, hacer reservas de restaurant y/o hospedaje. Permite el pago de entradas, acceso a realidad aumentada. Tiene acceso al clima, y permite acceder a un traductor.

En la aplicación “Itinerario” se pretende integrar varias tecnologías entre ellas:

- **SuperApps**

Las SuperApps son aplicaciones móviles que ofrecen una amplia gama de servicios y funciones en una sola plataforma, lo que las convierte en una opción conveniente para los usuarios que buscan simplificar sus vidas digitales.

- **Serverless**

Es una arquitectura de computación en la nube que está ganando popularidad en el mundo de las aplicaciones móviles de nueva generación. En lugar de tener que gestionar y mantener infraestructuras complejas y costosas, el desarrollo de aplicaciones serverless permite que los desarrolladores se centren en escribir código de aplicación sin preocuparse por la infraestructura subyacente. Esto se logra mediante la externalización de la gestión de servidores a los proveedores de servicios en la nube, como Amazon Web Services (AWS).

AWS Amplify es un conjunto de herramientas y servicios que permite a los desarrolladores crear y escalar aplicaciones móviles sin tener que preocuparse por la infraestructura subyacente. Amplify proporciona una plataforma completa para la construcción de aplicaciones móviles que incluye funcionalidades como la autenticación de usuarios, el almacenamiento de datos, el análisis y la integración con otras herramientas y servicios de AWS. Además, Amplify es compatible con una amplia gama de tecnologías y marcos de trabajo, lo que lo hace flexible y escalable.

En resumen, la arquitectura serverless es una forma innovadora y rentable de construir aplicaciones móviles de nueva generación, y AWS Amplify es una solución completa que ofrece a los desarrolladores una manera fácil de construir, escalar y administrar aplicaciones móviles. Con el aumento de la demanda de aplicaciones móviles de alta calidad y con un

rendimiento rápido y escalable, la adopción de la arquitectura serverless y las herramientas de Amplify es una opción inteligente para los desarrolladores móviles que buscan impulsar sus aplicaciones al siguiente nivel.

• IA Generativa

Las aplicaciones de IA generativa son programas que se basan en un tipo de IA que pueden crear nuevos contenidos e ideas, como conversaciones, historias, imágenes, videos, código y música. Al igual que todas las aplicaciones de IA, las aplicaciones de IA generativa funcionan con modelos que se entrenan con anterioridad con grandes cantidades de datos y, por lo general, se denominan modelos fundacionales (FM).

Existe un crecimiento acelerado en las startups de IA generativa, y también en las startups que crean herramientas para simplificar la adopción de la IA generativa. Herramientas como LangChain, un marco de código abierto para desarrollar aplicaciones basadas en modelos del lenguaje, hacen que la IA generativa sea más accesible para una gama más amplia de organizaciones, lo que permitirá una adopción más rápida. Estas herramientas también incluyen la ingeniería de peticiones (prompts), la ampliación de servicios (como la integración de herramientas o bases de datos vectoriales), la supervisión de modelos, la medición de la calidad de los modelos, las barreras de protección, la anotación de datos, el aprendizaje por refuerzo a partir de la retroalimentación humana (RLHF) y muchas más.

Los componentes comunes de una aplicación de IA generativa son el modelo fundacional (FM), su interfaz, opcionalmente la plataforma de machine learning (ML) y la computación acelerada. Estos se pueden cumplir con las ofertas administradas disponibles en la plataforma AWS, como Amazon Bedrock (modelo fundacional y sus componentes de interfaz).

Objetivos

El objetivo de este grupo de trabajo es explorar y analizar los beneficios de integrar a la SuperApp “Itinerario” con tecnologías avanzadas como machine learning, geolocalización e infraestructura serverless, con el fin de ofrecer recomendaciones dinámicas y adaptativas basadas en las preferencias y el comportamiento del usuario.

El estudio se enfocará en el impacto de la IA generativa para la personalización de sugerencias turísticas, pensado para mejorar la experiencia del usuario.

Es creciente la demanda de experiencias personalizadas, es decir, los viajeros modernos buscan experiencias adaptadas a sus gustos, presupuesto y preferencias, pero muchas plataformas actuales ofrecen recomendaciones genéricas y poco contextualizadas. Una SuperApp con IA permitiría personalizar la experiencia en tiempo real, mejorando la satisfacción del usuario.

Formación de Recursos Humanos

El equipo de trabajo de esta línea de investigación está compuesto de ocho investigadores de la universidad Nacional de San Juan. Además, el proyecto marco donde se está desarrollando esta propuesta ha establecido vínculos con investigadores de la Nacional de San Luis, de la Universidad Champagnat y de la Universidad Nacional de Salta y dos alumnos de grado más.
Se está desarrollando una tesis doctoral sobre paralelismo híbrido y Big Data, una tesis de maestría en áreas afines y dos tesinas de grado en el área de Serverless computing, Concurrencia y Computación distribuida.

Además, se espera aumentar el número de publicaciones. Por otro lado, también se prevé la divulgación de varios temas investigados por medio de cursos de postgrado y actualización o publicaciones de divulgación y asesoramiento a empresas y otras instituciones públicas y privadas.

Referencias

[1] Jonas E, Schleier-Smith J, Sreekanti V, Tsai C-C, Khandelwal A, Pu Q, Shankar V, Carreira J, Krauth K, Yadwadkar N, Gonzalez JE, Popa RA, Stoica I, Patterson DA (2019) Cloud Programming Simplified: A Berkeley View on Serverless Computing. http://arxiv.org/abs/1902.03383. Accessed 6 Jan 2021

[2] S. Eismann et al., 'Serverless Applications: Why, When, and How?' (2021), IEEE Software, vol. 38, no. 1, pp. 32–39, Jan. 2021, doi: 10.1109/MS.2020.3023302. problems. In Research Advances in Cloud Computing (pp. 1-20). Springer, Singapore.

[3] Nastic, S., Rausch, T., Scekic, O., Dustdar, S., Gusev, M., Koteska, B. & Prodan, R. (2017). A serverless real-time data analytics platform for edge computing. IEEE Internet Computing, 21(4), 64-71.

[4] Mohanty, S. K., Premsankar, G., & Di Francesco, M. (2018). An Evaluation of Open Source Serverless Computing Frameworks. In CloudCom (pp. 115-120).

[5] Gottlieb, N. (2016). State of the Serverless Community Survey Results.https: //serverless.com/blog/state-of-serverlesscommunity/.

[6] Jonas, E., Pu, Q., Venkataraman, S., Stoica, I., & Recht, B. (2017). Occupy the cloud: Distributed computing for the 99%. In Proceedings of the 2017 Symposium on Cloud Computing (pp. 445-451). ACM.

[7] Fromm, K. (2012).https://readwrite.com/2012/10/15/why-the-future-of-software-and-apps-is-serverless/

[8] Van Eyk, E., Iosup, A., Seif, S., & Thömmes, M. (2017). The SPEC cloud group's research vision on FaaS and serverless architectures. In Proceedings of the 2nd International Workshop on Serverless Computing (pp. 1-4). ACM.

[9]https://es.wikipedia.org/wiki/Tecnologia_inmersiva

[10] Desarrollo de pila completa: AWS Amplify. https://aws.amazon.com/es/amplify/

[11] https://blog.back4app.com/amazonmbaas/

[12] https://backendless.com/backendless-vsaws-amplify/

Propuesta de Expansión Regional y Blockchain para la Transparencia en Datos Abiertos

Roxana Martínez, Horacio Caballero, Bruno Navarrete, Andrés Cheroni

Secretaría de Investigación y Transferencia Científica
Universidad Siglo 21 (S21)
Calle De los Latinos 8555, Córdoba, Argentina

{roxana.martinez, horacio.caballero}@ues21.edu.ar
{bruno.nvte, andrescheroni}@gmail.com

RESUMEN

Uno de los ejes básicos de la transparencia gubernamental, es la utilización de los datos abiertos. Estos son fundamentales para fortalecer la confianza ciudadana y mejorar la eficiencia de la gestión pública. Este proyecto propone una doble estrategia: la comparación internacional de buenas prácticas en datos abiertos y la exploración del uso de Blockchain como tecnología para garantizar la trazabilidad e inmutabilidad de la información pública. A través del análisis de los modelos implementados en Argentina, y algunos países de Latinoamérica, se identificarán fortalezas y áreas de mejora en la gestión de datos abiertos. Además, se evaluará el impacto de Blockchain en la transparencia, proponiendo un modelo conceptual para su aplicación en entornos gubernamentales. Se espera que los resultados permitan establecer recomendaciones para optimizar la accesibilidad, fiabilidad y uso de los datos públicos.

Por otra parte, existen algunas vulnerabilidades en el tratamiento, ya que, en muchos casos, existen datos sensibles, como ser información personal, y la falta de medidas de seguridad adecuadas podría resultar en generar brechas de seguridad y violaciones de la privacidad. Otro aspecto fundamental en esta situación es la calidad de datos del contenido disponibilizado, debido a que esto puede afectar la eficacia de las políticas y decisiones basadas en estos datos. Finalmente, es necesario intervenir en criterios de coordinación y estandarización, ya que varias entidades gubernamentales recopilan datos similares y si estos no se encuentran bien gestionados en su tratamiento, podrían existir tareas redundantes. Como resultados deseados, se espera obtener el estado y la situación sobre el tratamiento de los datos públicos en el contexto de transparencia para el paradigma Gobierno Abierto.

Palabras clave: Gobierno Abierto, Transparencia, Datos públicos, Blockchain, Comparación Internacional.

CONTEXTO

El presente trabajo es parte del proyecto denominado "*Tratamiento de Datos Públicos para fomentar la transparencia en Gobierno Abierto*", que tuvo inicio en el mes de diciembre 2023. Este proyecto pertenece a la línea temática de investigación de Transformación Tecnológica, del área de investigación de Ciencias Aplicadas, por parte de la Secretaría de Investigación y Transferencia Científica. Las actividades del proyecto están directamente relacionadas con las carreras grado del Área de Ingeniería y Sistemas de la Universidad Siglo 21 (S21). El proyecto es financiado y evaluado por la Secretaría de Investigación y Transferencia Científica. Además, cuenta con la participación de docentes y estudiantes de grado de la carrera de Licenciatura en Informática. Una de las líneas a presentar del proyecto de investigación es el titulado "*Propuesta de Expansión Regional y Blockchain para la Transparencia en Datos Abiertos*", el cual se desarrolla como una continuidad y evolución del proyecto previo

"*Tratamiento de Datos Públicos para fomentar la transparencia en Gobierno Abierto*". El objetivo central es analizar el estado actual de la gestión de datos abiertos en el ámbito gubernamental, evaluando estrategias de expansión regional e incorporando Blockchain como tecnología para fortalecer la trazabilidad e inmutabilidad de la información pública. La investigación se orienta a identificar fortalezas y desafíos en los modelos implementados en Argentina y otros países de Latinoamérica, promoviendo una visión comparativa de buenas prácticas internacionales.

1. INTRODUCCIÓN

En los últimos años, la transparencia gubernamental y la apertura de datos han tomado un papel central en la agenda de transformación digital de los gobiernos. La implementación de políticas de datos abiertos busca promover el acceso a la información pública, fortalecer la confianza ciudadana y mejorar la eficiencia en la gestión gubernamental [1]. Autores como Rojas [2] han explorado la integración de herramientas de inteligencia de negocios en portales de datos abiertos, utilizando Lenguajes de Dominio Específico (DSL) para facilitar la estructuración y consulta de la información en entornos de Gobierno Abierto. A pesar de los avances en la adopción de datos abiertos, persisten desafíos relacionados con la calidad, accesibilidad y estandarización de la información publicada [3]. Cabe mencionar que trabajos de investigaciones previas han identificado la necesidad de contar con criterios claros de publicación, que incluyan vocabularios normalizados y enfoques metodológicos que favorezcan la interoperabilidad de los datos [4]. En este sentido, estudios como el de Venegas Álvarez [5] han evaluado la reutilización de datos abiertos en distintos países de Latinoamérica, resaltando la importancia de estrategias orientadas a maximizar su impacto y utilidad. Uno de los desafíos clave es la falta de actualización periódica de los conjuntos de datos y la ausencia de métricas de calidad que permitan evaluar su confiabilidad y aplicabilidad en la toma de decisiones [7]. Además, la descentralización en la gestión de datos abiertos puede derivar en esfuerzos duplicados y falta de coordinación entre las distintas entidades gubernamentales. Para abordar estos problemas, investigaciones recientes han propuesto la incorporación de tecnologías como Blockchain [8] [9], que permiten garantizar la trazabilidad e inmutabilidad de los datos públicos, mejorando su transparencia y seguridad [10].

La transparencia en los datos públicos es un elemento fundamental para la gobernanza moderna. La ciudadanía exige cada vez más acceso a información confiable y verificable, lo que impulsa la necesidad de fortalecer los mecanismos de integridad en la publicación y manejo de estos datos [11] [12]. En este contexto, Blockchain emerge como una tecnología innovadora que permite registrar transacciones de manera descentralizada, asegurando que la información no pueda ser alterada ni manipulada sin dejar rastro [13]. Dado que los datos abiertos pueden ser utilizados para la fiscalización de políticas públicas, la incorporación de Blockchain posibilita una auditoría constante de la información [14]. Esto es especialmente relevante en áreas como la asignación de presupuestos gubernamentales, contrataciones públicas y ejecución de proyectos, donde la falta de trazabilidad ha sido históricamente un problema. Al aplicar esta tecnología, se fortalece la transparencia institucional y se minimiza el riesgo de corrupción, ya que cualquier intento de modificación indebida en los registros será inmediatamente detectable [15]. Otro aspecto clave es la capacidad de Blockchain para mejorar la interoperabilidad de los datos abiertos. A través de contratos inteligentes y mecanismos de certificación digital, es posible garantizar la autenticidad de los datos en diferentes plataformas gubernamentales. Esto no solo facilita el acceso de los ciudadanos a la información, sino que también optimiza la cooperación entre organismos estatales, evitando la duplicación de esfuerzos en la recolección y almacenamiento de datos [16].

Este trabajo tiene como objetivo analizar la gestión de datos abiertos en Argentina y otros países de la región, identificando buenas prácticas y oportunidades de mejora en términos de calidad, accesibilidad y seguridad. Asimismo, se explorará el impacto de Blockchain como una herramienta clave para optimizar la transparencia en la publicación y gestión de datos gubernamentales. A partir de estos análisis, se presentará un modelo conceptual para su implementación, con el propósito de contribuir al fortalecimiento del paradigma de Gobierno Abierto en la región.

2. LÍNEAS DE INVESTIGACIÓN DESARROLLO

Este proyecto pertenece a la línea temática de investigación de Transformación Tecnológica, y al área de Investigación de Ciencias Aplicadas de la Secretaría de Investigación y Transferencia Científica. Los ejes principales del tema que se están investigando en referencia a las actividades de I+D+i (Investigación, Desarrollo e Innovación) son:

- Relevar y comparar estudios previos sobre la implementación de datos abiertos en gobiernos de la región y Blockchain.
- Analizar modelos de aplicación de Blockchain en la transparencia de datos públicos y su impacto en la trazabilidad e inmutabilidad de la información gubernamental.
- Diseñar recursos didácticos que puedan ser utilizados por los estudiantes de las carreras involucradas en el proyecto.
- Evaluar vulnerabilidades en la gestión de datos abiertos, particularmente en la seguridad y privacidad de la información.
- Proponer estrategias de estandarización y coordinación en la recopilación y gestión de datos gubernamentales para minimizar redundancias.

Planteamiento de la problemática:
En el contexto actual, el tratamiento de datos públicos enfrenta varios desafíos relacionados con la transparencia, la seguridad y la interoperabilidad de la información gubernamental. A pesar de los avances en la apertura de datos, muchas iniciativas presentan limitaciones en cuanto a su accesibilidad, calidad y confiabilidad. La falta de estandarización en la recopilación y publicación de datos dificulta su comparación y reutilización por parte de los ciudadanos y organismos de control. Uno de los principales obstáculos es la desconfianza de la población respecto a la manipulación y uso de los datos públicos. La ausencia de mecanismos de afecta la credibilidad de la información publicada. En este sentido, Blockchain emerge como una tecnología potencial para fortalecer la transparencia y trazabilidad de los datos, permitiendo registros inmutables y accesibles. Por otra parte, la exposición de datos sensibles sin medidas de seguridad adecuadas representa un riesgo significativo en términos de privacidad y posibles brechas de seguridad.

Pregunta de investigación del proyecto: ¿Cómo pueden la expansión regional y el uso de Blockchain mejorar la transparencia y la gestión de datos públicos en el marco del Gobierno Abierto?

Hipótesis:
La integración de Blockchain en la gestión de datos abiertos fortalece la transparencia y la confianza ciudadana, al garantizar la inmutabilidad y trazabilidad de la información pública. Además, la implementación de políticas de estandarización y coordinación entre entidades gubernamentales reduce redundancias y mejora la accesibilidad y calidad de los datos públicos.

3. RESULTADOS OBTENIDOS/ESPERADOS

Enfoque del proyecto:
El presente proyecto se orienta a evaluar y mejorar la transparencia en la gestión de datos públicos mediante la implementación de tecnologías innovadoras, como Blockchain. Como parte de los resultados esperados, se busca fortalecer las políticas de datos abiertos, asegurando su trazabilidad, inmutabilidad y accesibilidad para la ciudadanía. Además, este

enfoque se vincula estrechamente con las áreas de Ingeniería de Software y Arquitectura de Software, proponiendo algunas soluciones tecnológicas concretas para optimizar la disponibilidad y el uso de los datos gubernamentales.

Objetivos principales:
Brindar buenas prácticas en la utilización de datos públicos para promover una adecuada apertura y accesibilidad de la información gubernamental.

Objetivos específicos esperados:
- Ampliar el análisis de buenas prácticas en la gestión de datos abiertos a nivel regional.
- Evaluar el impacto de la tecnología Blockchain en la trazabilidad y seguridad de los datos gubernamentales.
- Diseñar un modelo conceptual para la implementación de Blockchain en la gestión de datos abiertos.
- Desarrollar prototipos de software.
- Proponer estrategias para la coordinación y estandarización en la recopilación y gestión de datos públicos.
- Diseñar recursos de apoyo para la implementación de las buenas prácticas.

Metodología y Técnicas:
Para alcanzar estos resultados, se sigue una metodología mixta que combina técnicas cualitativas y cuantitativas. Se realiza un relevamiento documental de fuentes gubernamentales y estudios previos, complementado con análisis comparativos entre distintas plataformas de datos abiertos en la región. Además, se estudian casos específicos en los que Blockchain ha sido aplicado a la gestión pública, con el fin de evaluar su viabilidad en el contexto latinoamericano. Finalmente, se prevé la elaboración de un prototipo de validación de datos basado en Blockchain, que permitirá evaluar su impacto en términos de transparencia, seguridad y accesibilidad de la información pública. Por último, se considerará la evaluación de las limitaciones del estudio, entre las cuales se incluyen posibles desafíos tecnológicos y de infraestructura que podrían afectar el acceso a determinados archivos de datos públicos, dada su dependencia de la disponibilidad y gestión por parte de los organismos estatales.

Resultados obtenidos:
Como parte de los avances en esta línea de investigación, se han producido publicaciones en congresos científicos que reflejan el impacto del proyecto en el análisis de transparencia gubernamental en Argentina. Entre estos resultados, se destacan los siguientes estudios: “Iniciativas Tecnológicas para la Transparencia en Argentina: Un Relevamiento de Herramientas Gubernamentales” [17]: Este trabajo presenta un relevamiento de diversas iniciativas gubernamentales enfocadas en la transparencia y el acceso a la información pública. Se analizaron herramientas como el chatbot Tina, el portal de datos abiertos Andino y la plataforma Consulta Pública, evaluando su funcionalidad, accesibilidad y eficacia en la rendición de cuentas. Si bien se identificaron progresos, también se señalaron desafíos en términos de escalabilidad y eficiencia de estas plataformas. Por otro lado, el trabajo de investigación basado en la “Evaluación de la Transparencia en el Gobierno Abierto: Estudio de Caso en Argentina” [18]: En este estudio se realizó un diagnóstico sobre la disponibilidad y calidad de los datos abiertos en distintos organismos gubernamentales de Argentina. Se identificaron fortalezas en la implementación de la Ley de Acceso a la Información Pública, pero también falencias en la uniformidad y estandarización de los datos, lo que dificulta su reutilización y comparación a nivel nacional.

4. FORMACIÓN DE RECURSOS HUMANOS

Este proyecto se compone por 2 (dos) docentes de grado que cuentan con varios años de experiencia en el dictado de clases en varias Universidades de Argentina y más de 20 años de experiencia en el ambiente laboral en Tecnología Informática, también tienen estudios de posgrado: Uno de ellos es docente

investigadora, Doctora en Ciencias Informáticas de la Universidad Nacional de La Plata (UNLP), actual directora del Proyecto de Investigación y Magister en Tecnología Informática de la Universidad Abierta Interamericana (UAI), además, el proyecto cuenta con otro docente investigador, Magister en Tecnología Educativa de la Universidad Abierta Interamericana (UAI). El equipo también cuenta con la participación de estudiantes y egresados de grado de las carreras de la rama de Ingeniería y Sistemas de la Universidad Siglo 21, puntualmente de la carrera Licenciatura en Informática.

5. BIBLIOGRAFÍA

[1] García García, J. (2014). Gobierno abierto: transparencia, participación y colaboración en Administraciones Públicas. *Innovar*, *24*(54), 75-88.

[2] Rodríguez Rojas, L. A. (2017). Metamodelo para integración de datos abiertos aplicado a inteligencia de negocios. Tesis de Doctoral, Repositorio Institucional de la Universidad de Oviedo

[3] Martínez, M. R. (2022). Métricas de calidad para validar los conjuntos de datos abiertos públicos gubernamentales (Doctoral dissertation, Universidad Nacional de La Plata).

[4] Ruiz-Cueva, J., Pinedo-Tuanama, L., & Barbarán-Mozo, P. (2022). Gestión del Gobierno Abierto y uso del portal de transparencia en una universidad pública peruana. *Enfoque UTE*, *13*(1), 73-81.

[5] Venegas Álvarez, M. A. (2018). Estudio sobre el acceso abierto a datos gubernamentales: El caso de transparencia en Chile, Costa Rica y Uruguay. Tesis de Maestría, Universidad De Chile, Facultad de Ciencias Físicas y Matemáticas.

[6] Martínez, R. (2022). Tipos de Métricas de calidad para validar datasets gubernamentales argentinos. *Revista Abierta de Informática Aplicada*, *6*(2), 1-14.

[7] Pinho De Oliveira, M. F. (2022). Gobierno abierto: transparencia gubernamental en el marco de la pandemia covid-19. *Derecho global. Estudios sobre derecho y justicia*, *7*(20), 147-179.

[8] Davila, W. A. (2024). Blockchain la oportunidad de transformación digital en la administración pública del órgano judicial de Bolivia. *Investigación & Negocios*, *17*(30), 28-33.

[9] Gómez Reategui, J. F., & Chamoli Falcón, A. W. (2024). Revisión de la literatura sobre blockchain y confianza institucional: Perspectivas desde América Latina. *Aula Virtual*, *5*(12).

[10] Martínez, M. R., & Pincay-Ponce, J. I. (2024). Buenas Prácticas de Seguridad para la Protección de la Privacidad con Datos Abiertos. *Revista Científica de Informática ENCRIPTAR-ISSN: 2737-6389.*, *7*(14), 187-205.

[11] Morocho-Marín, J. P., & Montero-Cobo, M. A. (2024). Fortaleciendo la transparencia y la ética en la contabilidad pública: desafíos y perspectivas en América Latina. *Revista Metropolitana de Ciencias Aplicadas*, *7*(2), 30-41.

[12] Martínez, R., Rodríguez & et al. (2020). Análisis del procedimiento para la solicitud de información pública y tiempos de respuesta. In XIV Simposio de Informática en el Estado (SIE 2020)-JAIIO 49 (Modalidad virtual).

[13] Benítez Martínez, F. L. (2021). Un modelo de gobernabilidad para procesos de edemocracia basados en una red neuronal de blockchain.

[14] Mauras Vélez, A. L. (2022). *Verificador de integridad de documentos con tecnología blockchain para la Universidad Católica de Córdoba* (Doctoral dissertation, Universidad Católica de Córdoba).

[15] Estrada Beltran, A. G., & Cabezas Naranjo, M. W. (2024). *Implementación del Blockchain y contratos inteligentes en la contratación pública entre el GAD de Guano y actores privados: Retos y oportunidades* (Bachelor's thesis, Riobamba).

[16] Vicente-Paños, A., & Jordán-Alfonso, A. (2017). Acceso a la información pública y su reutilización en las comunidades autónomas: evaluación de la reutilización de datos abiertos. Profesional de la Información, 26(3), 381-391.

[17] Martínez, R., Navarrete B. et al. (2024). *Evaluación de la Transparencia en el Gobierno Abierto: Estudio de Caso en Argentina*. Congreso Argentino de Ciencias de la Computación (CACIC).

[18] Martínez, R., Cheroni A. et al. (2024). *Iniciativas Tecnológicas para la Transparencia en Argentina: Un Relevamiento de Herramientas Gubernamentales*. Congreso Nacional de Ingeniería Informática – Sistemas de Información (CoNaIISI). Universidad Nacional de Catamarca. Argentina.

SISTEMA DE ALERTAS DE GEOFENCING PARA EL MONITOREO DE ADULTOS MAYORES BASADOS EN ANDROID Y WEAR OS

Esteban A. Carnuccio, Pablo M. Vera, Rocío A. Rodríguez

Centro de Altos Estudios en Tecnología Informática (CAETI)
Facultad de Tecnología Informática
Universidad Abierta Interamericana (UAI)
Montes de Oca 745, Ciudad Autónoma de Buenos Aires, Argentina

{estebanandres.carnuccio} @alumnos.uai.edu.ar
{pablomartin.vera, rocioandrea.rodriguez} @uai.edu.ar

RESUMEN

Con el paso de los años, las personas adultas mayores requieren cada vez más cuidados y control por parte de familiares u otras personas, ya que pueden presentar dificultades en su vida cotidiana. A pesar de ello, es fundamental preservar su independencia y libertad en sus acciones diarias, para que no se sientan invadidos en su privacidad. En este sentido, los smartphones y los smartwatches ayudan en el monitoreo de estas personas. Es importante que estos dispositivos sean fáciles de usar para los adultos mayores, manteniendo su usabilidad. En este contexto, este artículo presenta los avances en el desarrollo de un sistema de geofencing para adultos mayores, utilizando teléfonos inteligentes Android y un smartwatch con Wear OS. El enfoque principal está en la adaptación de las notificaciones en el smartwatch para mejorar su accesibilidad junto con las notificaciones automáticas disparadas por ubicación. Se abordan las funcionalidades claves en desarrollo y los desafíos actuales en su optimización, con el objetivo de ofrecer una alternativa de monitoreo no invasivo para este grupo de personas.

Palabras clave: Android, IoT, Geolocalización, Geofencing, Weareable, Wear Os

CONTEXTO

Esta línea de investigación y desarrollo (I+D) forma parte de los proyectos de la Facultad de Tecnología Informática radicados en el Centro de Altos Estudios en Tecnología Informática (CAETI) sede Ciudad Autónoma de Buenos Aires de la Universidad Abierta Interamericana (UAI). El proyecto cuenta con financiamiento asignado.

1. INTRODUCCIÓN

A medida que pasan los años, una persona va reduciendo determinadas capacidades, tanto físicas como mentales o psicomotrices, lo que se acentúa en las personas mayores de 65 años. A partir de esa edad, según la OMS (Organización Mundial de la Salud) [1], son consideradas personas mayores. En este sentido, en algunos casos, esto termina generando una preocupación constante en los familiares de los adultos mayores, dado que pueden sentirse intranquilos al no saber si tienen algún inconveniente en su vida diaria. Una de las características que presentan las personas de avanzada edad son los olvidos de eventos o la desorientación en determinados lugares. Un caso típico puede ser, por ejemplo, olvidarse de asistir al médico en un día determinado. Otra de las preocupaciones clave que pueden tener los familiares es si el adulto mayor llegó bien a su hogar después de

regresar de un determinado sitio, como volver a su casa tras un cumpleaños.

Por otro lado, si los adultos mayores se sienten muy controlados en su vida diaria, esto puede llegar a afectar su estado de ánimo, ya que sentirán que pierden su independencia y privacidad [2]. Por ese motivo, una posible herramienta que podría ayudar a esta problemática es el uso de dispositivos wearables (vestibles), como puede ser un smartwatch o incluso sus propios smartphones. De este modo poseen elementos que no son invasivos y les permiten a las personas mayores continuar con su vida cotidiana sin sentirse restringidos. Hoy en día, la gran mayoría de las personas usa un teléfono o un reloj inteligente, siendo estos los más utilizados [3] [4], ya sean por niños, adolescentes, adultos o personas mayores. Es en este último grupo etario en el que se centra esta investigación.

Actualmente existen muchos dispositivos de este tipo. Los más utilizados son los smartphones, que generalmente tienen instalado el sistema operativo Android. A pesar del crecimiento de Apple en el mercado de dispositivos móviles, Android sigue siendo el sistema operativo líder a nivel mundial. En 2024, el 71,42% de los teléfonos móviles utilizaban Android, mientras que iOS alcanzaba una cuota de mercado cercana al 28% [5]. Por otro lado, Wear OS es uno de los sistemas más utilizados en smartwatches, dado que es una implementación de Android y viene instalada en diversos equipos.

En la Tabla 1 se presentan sistemas operativos que están integrados en smartwatch comerciales, siendo los más populares, cabe destacar que existen otros sistemas operativos pero que se encuentran en un solo dispositivo en particular, no siendo de uso masivo.

Tabla 1. Sistemas Operativos para principales marcas de smartwaches

Sistema Operativo	**Utilizado en**
Wear OS	Android
watchOS	Apple
Tizen	En algunos Samsumg

Mediante Geofencing, es posible monitorear actividades de interés de una persona sin que se sienta controlada. Esta herramienta les permite a los familiares del adulto mayor ser alertadas cuando el individuo ingresa o sale de determinado perímetro geográfico. Lo que evita que el familiar este constantemente, consultándole al adulto mayor su ubicación geográfica, ya que esta será alertada de manera automática, cuando sucede el evento [6].

Existen diferentes metodologías para utilizar geofencing, la más utilizada es emplear servicios externos en la nube para realizar el procesamiento geográfico. Pero esto encarecería el servicio dado los costos operativos diarios como el uso de servidores en la nube, almacenamiento y ancho de banda. Por ese motivo, en esta investigación se aplica la técnica de geofencing en el propio teléfono inteligente. De esta forma, no hay costos adicionales. Además, de esta manera se mejora la privacidad y la seguridad de los datos ya que no hay transmisión de datos sensibles con servidores externos, protegiendo la privacidad del usuario [7].

Un factor bastante importante que se debe tener en cuenta en el diseño de los dispositivos es que sean fáciles de usar para las personas mayores, ya que deben ser accesibles para ellas. La mayoría de los smartwatches presentan mensajes complejos e interfaces poco amigables para este grupo, lo cual se detalla en la parte de resultados

con respecto a la adaptación que fue necesaria hacer sobre las notificaciones. Una interfaz simple y amigable, influye directamente en que el adulto continúe utilizando la aplicación.

Por este motivo, en este artículo se presentan los avances en el desarrollo de un sistema de Geofencing compuesto por dos dispositivos: un smartphone Android y un smartwatch con Wear OS, destinados al uso de adultos mayores. Es importante mencionar que el sistema aún se encuentra en desarrollo, con funcionalidades clave en implementación y en proceso de optimización.

2. LINEAS DE INVESTIGACIÓN Y DESARROLLO

Este proyecto de investigación busca:

- Hacer uso de la geolocalización y de las técnicas de Geofencing para aplicarlos a aplicaciones Wearables.
- Diseñar una aplicación de monitoreo no invasiva en Android y Wear OS.
- Realizar el sistema de Geofencing configurable por parte del usuario
- Realizar un sistema con una interfaz amigable y de uso rápido por parte de las personas mayores de edad

3. RESULTADOS OBTENIDOS/ESPERADOS

El sistema se está desarrollando, utilizando el Framework que ofrece Android Studio, a través del lenguaje de programación Kotlin. En su implementación se aplican dos patrones de diseño: por un lado, MVC (Model View Controller) y, por otro, MVVM (Model -View-ViewModel), ambos ampliamente utilizados con este lenguaje.

Dado que el smartwatch tiene una capacidad de energía limitada, no es recomendable realizar el procesamiento de Geofencing en este dispositivo. Esto se debe a que esta técnica requiere un consumo energético que el reloj no puede sostener a lo largo del tiempo. Por este motivo, el sistema ha sido diseñado para que la activación de los eventos de Geofencing sea procesada en el smartphone en tiempo real, ya que este dispone de mayor capacidad energética. De esta manera, el teléfono se encarga del control del área de Geofencing, mientras que el smartwatch alerta al usuario de manera sencilla y amigable en caso de que ocurra algún evento relacionado con la geocerca. Además, el smartwatch detecta de forma limitada eventos de caída y, cuando se produce uno de estos eventos, se envía un SMS al familiar responsable.

El sistema permite generar múltiples áreas de Geofencing, las cuales pueden ser creadas de forma gráfica mediante las APIs que provee Google Maps. Para su activación, se utiliza el servicio de Geofencing proporcionado por el Framework de Android. Estas áreas pueden configurarse para generar alertas cuando la persona ingresa en la zona definida, cuando sale de ella o cuando permanece dentro durante un tiempo determinado.

Fue necesario adaptar las notificaciones para Wear OS para mejorar la usabilidad de los adultos mayores, de manera que sean fáciles y sencillas de utilizar. Esto se debe a que las notificaciones predeterminadas en Wear OS pueden resultar complejas para las personas mayores, ya que agrupan diferentes avisos en el Stream de notificaciones del reloj. Por ejemplo, en la misma bandeja de mensajes pueden aparecer notificaciones de WhatsApp, SMS, Facebook, entre otras aplicaciones. Además, suelen incluir demasiadas opciones de botones dentro de un mismo mensaje, lo que dificulta su

uso para un adulto mayor. Esto se puede visualizar en imagen izquierda de la Fig. 1, en donde se muestra cómo se presenta a un adulto mayor una notificación de detección de caída, la cual está agrupada en el Stream de notificaciones del sistema operativo, junto con otros mensajes que no corresponden a la aplicación. Como resultado, si el adulto mayor desea enviar un SMS de pedido de ayuda a un familiar, primero deberá desplegar la notificación y luego pulsar el botón de SOS, tal como se muestra a la derecha. Este proceso puede resultar complicado y poco amigable para una persona mayor.



Fig. 1. Stream de Notificaciones de Wear OS

Por ese motivo, se determinó conveniente desarrollar una aplicación de notificaciones personalizadas que muestran de manera más amena únicamente las notificaciones generadas por el sistema de monitoreo de adultos mayores, todas agrupadas en la misma interfaz gráfica. En donde se muestra solamente la información pertinente. Esto se puede visualizar en la Fig. . Como se puede observar, las notificaciones están agrupadas en páginas, mostrando siempre la última notificación activa. Por lo tanto, si el usuario desea visualizar los mensajes anteriores, deberá deslizar la pantalla hacia los costados. En consecuencia, el indicador de paginación mostrará en qué notificación se encuentra. Luego, si el usuario desea enviar el SMS de ayuda, deberá presionar el símbolo de color rojo que indica la alerta.



Fig. 2. Notificaciones adaptadas para Wear OS de Detección de Caídas para adultos mayores

Es importante mencionar que el envío de SMS se realiza desde el smartphone. Por lo que, el smartwatch se comunica a través de bluetooth con el teléfono para solicitar el envío del mensaje de texto. A su vez, para la detección de caídas con el smartwatch, se ha utilizado el servicio de sensores de Android Studio, que permite detectar estos eventos mediante algoritmos.

En la Fig. 2, se puede visualizar la generación de áreas de geofencing de forma visual. En donde cada una de ellas pueden presentar diferentes tamaños de cobertura de activación. A su vez se puede observar la generación de una notificación de alerta de un área de geofencing, que se genera en este caso en los dos dispositivos que usará el adulto mayor. Por un lado, en el smartphone y por otro, en el propio smartwatch. En este último generando una alerta de atención de que ingreso a una determinada área.

Otro aspecto que también se puede observar en la Fig. 2 es que este tipo de notificaciones adaptadas podrían utilizarse para generar recordatorios de eventos. Por ejemplo, un aviso de que el usuario debe asistir a un control médico en una fecha determinada. Si el usuario presiona el botón de calendario de color verde, se mostrará en el mapa la ubicación del lugar especificado por una geocerca. Este tipo de notificaciones de recordatorio se pretende desarrollar en futuros trabajos, ya que aún no se encuentra implementado.

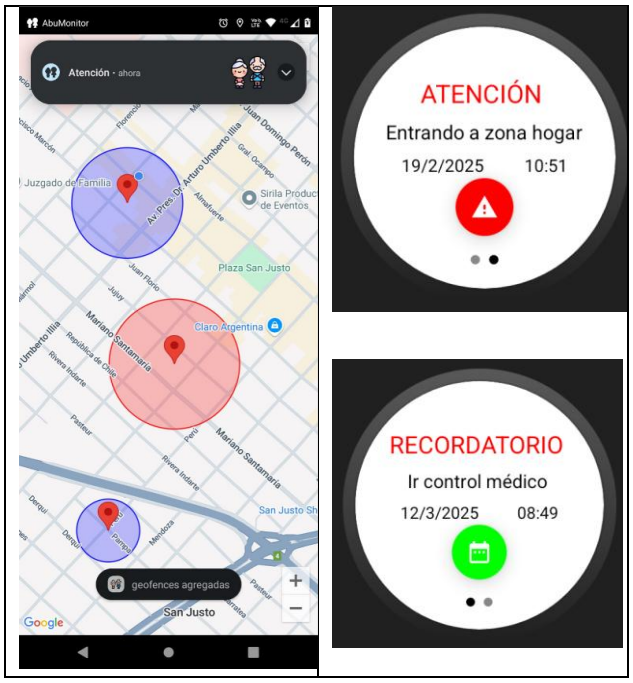


Fig. 2 .Notificaciones adaptadas de Geofencing para adultos mayores en Wear OS

En desarrollos futuros, se espera que estas áreas también se puedan configurar para activarse en una fecha y horario determinado. De este modo, si la persona no ingresa en un sitio dentro del tiempo estipulado, se emitirá una alerta al smartwatch del adulto mayor, además de notificar a sus familiares. También se planea asignar diferentes niveles de prioridad a las áreas, permitiendo que el usuario defina si desea o no ser alertado según la prioridad establecida en cada una.

4. FORMACIÓN DE RECURSOS HUMANOS

El grupo está formado por 5 personas, docentes de grado, postgrado y alumnos.

En el área de dispositivos móviles se encuentran en realización, 2 tesis de maestría y 1 tesina de grado en la UAI (Universidad Abierta Interamericana).

5. BIBLIOGRAFÍA

[1] O. M. d. l. Salud, Informe mundial sobre el envejecimiento y la salud. Organización Mundial de la Salud., 2015.

[2] J. Ardila, A. Ardila, L. Orejarena y L. Silva, «Relación del nivel de dependencia y la depresión en los adultos mayores de un hogar vida de Floridablanca, Santander en el año 2018,» Universidad Cooperativa de Colombia, 2018.

[3] Ditrendia, «Estadisticas sobre móviles en el mundo,» 2025. https://mktefa.ditrendia.es/blog/estadisticas-sobre-m%C3%B3viles-2023.

[4] Statista, « Smart wearable shipments forecast worldwide from 2016 to 2026,» 2025. https://www.statista.com/statistics/878144/worldwide-smart-wristwear-shipments-forecast/.

[5] Statista, «Cuota de mercado mundial de los sistemas operativos para dispositivos móviles de 2010 a 2024,» 2025. https://es.statista.com/estadisticas/635202/sistemas-operativos-de-telefonos-moviles-cuota-de-mercado-mundial/.

[6] K. Zuva, «Tracking of Customers using Geofencing Technology,» 2019.

[7] T. Ashley, «NO PLACE TO HIDE: Privacy Implications of Geolocation Tracking and Geofencing,» *Scitech Lawyer,* vol. 16, nº 2, pp. 20-23, 2020.

Tremün+: Plataforma para entrenamiento cognitivo mejorada

Lourdes Valli, Tomás Dozo, Lautaro Lopez, Laura Fava, Matás Pagano, Diego Vilches
LINTI - Laboratorio de Investigación en Nuevas Tecnologías Informáticas.
Facultad de Informática. Universidad Nacional de La Plata, Calle 50 esq. 120, 2do Piso.
Tel: +54 221 4223528
lourdes.valli@alu.ing.unlp.edu.ar, tomas.dozo@ing.unlp.edu.ar, {lfava, matiasp, dvilches}@linti.unlp.edu.ar, jdiaz@unlp.edu.ar

RESUMEN

La integración cognitiva y emocional en el entrenamiento deportivo conduce a mejorar el rendimiento y la toma de decisiones de los deportistas. Es por ello que los entrenadores y preparadores físicos siempre buscan formas de mejorar las habilidades visuales y visomotoras con programas de entrenamiento.

El eje de esta línea de I+D lo constituye la actualización de una plataforma para entrenamiento cognitivo destinados a mejorar la performance de deportistas que integra tecnologías y protocolos estándares de IoT.

Palabras claves: dispositivos para entrenamiento, tecnología en deporte, MQTT, Internet of Things (IoT).

CONTEXTO

El Laboratorio de Investigación de Nuevas Tecnologías Informáticas (LINTI) de la Facultad de Informática, viene trabajando en proyectos relacionados con tecnologías aplicadas al deporte acompañando las metodologías clásicas de entrenamiento, así como también, la administración de las nuevas fuentes de datos y mediciones existentes (Fava, L., 2018).

El uso de estos dispositivos y de las tecnologías de IoT nos ofrecen una potente fuente de información para planear el progreso del deportista y optimizar sus entrenamientos y competiciones (Alexandre D, 2012). Hoy en día, los profesionales del deporte pueden disponer de datos exactos y en tiempo real, con los que medir velocidad, distancias recorridas, movimientos realizados y una enorme cantidad de datos que abren una nueva línea de trabajo vinculada con el análisis de datos, aplicación de algoritmos de Machine Learning y la creación de modelos para realizar predicciones.

La línea de trabajo que se describe en este artículo se desarrollan en el LINTI y están enmarcadas en el proyecto de I+D *Tecnologías Digitales para la Inclusión, la Equidad y la Sostenibilidad*, acreditado en el marco del Programa de Incentivos, bajo la dirección del Lic. Javier Díaz.

1. INTRODUCCIÓN

El término Internet de las Cosas o Internet of Things (IoT) se refiere generalmente a escenarios donde la capacidad de cómputo y la conectividad de las redes se extienden a objetos, sensores y elementos cotidianos, permitiendo que estos dispositivos generen, intercambien y consuman datos. El concepto de combinar computadoras, sensores y redes para monitorear y controlar diferentes dispositivos ha existido durante décadas. Sin embargo, la reciente confluencia de

diferentes tendencias del mercado tecnológico está generando innovaciones en múltiples sectores.

Los nuevos productos de IoT para lograr "ciudades inteligentes" son una realidad y como era de esperar IoT también está alcanzando la industria del deporte. La utilización de dispositivos inteligentes o wearables y el análisis de datos que generan en tiempo real están cambiando el mundo del deporte, y los deportistas podrían optimizar su rendimiento y adecuar sus entrenamientos haciendo uso de estas tecnologías.

La ciencia está descubriendo que si bien toda la actividad física tiene un efecto positivo en el cerebro, el ejercicio que combina el uso de múltiples sentidos con movimientos corporales completos, desafía al cerebro a niveles más altos, porque requieren una función cognitiva más compleja para tomar decisiones y ejecutar habilidades. En consecuencia, la práctica de estas actividades mejora la toma de decisiones y el rendimiento físico, sensorial y neurológico de los deportistas (Lamberti, C., 2018).

Otro motivo relevante es la creciente demanda de entrenamiento cognitivo en el deporte. Aunque es una disciplina joven con pocas opciones y costos elevados, su desarrollo y estudio están en aumento (Walton C, 2018).

En este contexto se continúa trabajando para la mejora y actualización de la plataforma de entrenamiento.

2. Líneas de Investigación, Desarrollo e Innovación

Las líneas de investigación, desarrollo e innovación que se llevan a cabo en este proyecto están vinculadas al desarrollo de dispositivos basados en sensores y al desarrollo de aplicaciones que permitan mejorar el entrenamiento de los deportistas de alto rendimiento.

Los ejes principales de I+D+i son:

- Análisis de tecnologías de vanguardia para la construcción de dispositivos y software que mejore el entrenamiento y rendimiento de los jugadores.
- Implementación de mejoras y extensiones en los nodos que forman parte de la plataforma Tremün+.
- Diseño y construcción de nuevos tipos de dispositivos para ser usados en las prácticas deportivas (dispositivos con tecnologías basados en sensores para entrenamiento de arqueros, de jugadores de hockey, etc.).
- Análisis de datos recolectados mediante la utilización de técnicas de ciencias de datos para la identificación de patrones y toma de decisiones.

3. Resultados y Objetivos

Los avances presentados en esta línea están relacionados principalmente con software, esto es, se han re-implementado y estandarizado varios componentes. La plataforma de entrenamiento incluye varios dispositivos, o nodos, con matrices de LEDs y sensores de proximidad, una aplicación móvil que permite crear y ejecutar rutinas y un controlador que funciona como mediador de la comunicación entre la aplicación móvil y los nodos (Fig. 1).

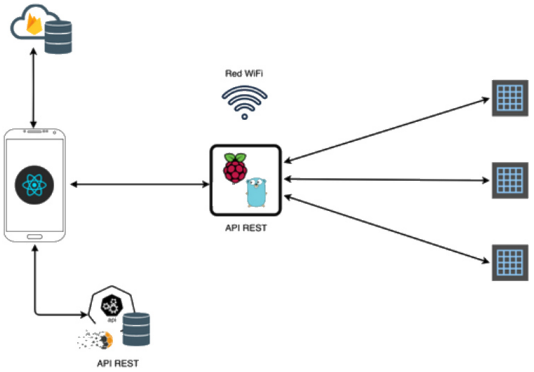


Fig 1: Arquitectura de la Plataforma

Se ha trabajado en mejorar los dispositivos y software asociado a ellos, migrando a nuevas tecnologías de comunicación y estandarizando el código existente. A continuación se describen los resultados de cada una de las líneas de I+D+i mencionadas en el inciso anterior:

Actualización del firmware de los nodos

La Fig. 2 muestra un nodo, donde se observa la placa de circuito impreso (PCB) de frente y dorso y el diseño de la carcasa.



Fig. 2: (a) Frente PCB (b)Dorso PCB (c) Carcasa

Los nodos cuentan con una matriz de LEDs independientes que permite armar figuras geométricas de diferentes colores y un sensor de proximidad para que el deportista al acercarse pueda apagarlo. El software del mismo fue migrado del sdk del fabricantes Arduino para estandarizar el proyecto y poder hacer uso de las librerías disponibles..

Desarrollo de una app de administración

Se implementó una app móvil para Android y iOS que permite la ejecución de rutinas preexistentes y la creación de nuevas y brinda retroalimentación sobre el desempeño de los deportistas mediante diferentes parámetros (velocidad, tiempo de reacción, etc). que pueden ser vistas al instante o almacenarse para posterior análisis. La Fig. 3 muestra capturas de pantalla de la misma. En la de la izquierda se pueden ver rutinas precargadas en la app y a la derecha se un wizard para crear rutinas personalizadas.

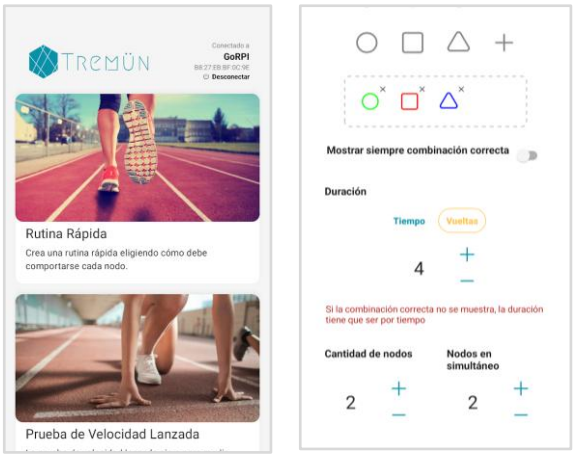


Fig. 3: Generación de rutina

Reemplazo de protocolos de comunicación

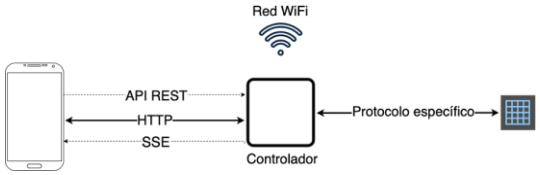
El protocolo MQTT (*Message Queuing Telemetry Transport*) es un protocolo de comunicación ligero basado en el modelo *publicación/suscripción*, diseñado para dispositivos con recursos limitados y redes con ancho de banda reducido o alta latencia. (A. S. Suwardi Ansyah, 2023).

Los dispositivos integrados con LEDs se comunican mediante el protocolo MQTT, a diferencia de la versión anterior que utilizaba un protocolo desarrollado específicamente para esta plataforma. Esta decisión se basó en la amplia

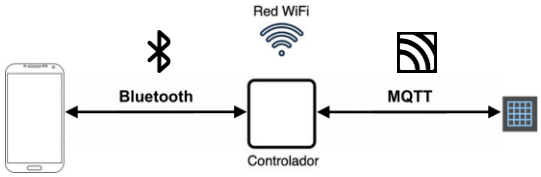
utilización de MQTT en la industria IoT y la posibilidad de estandarizar el proyecto para futuras mejoras.

La aplicación móvil desarrollada se comunica actualmente utilizando BLE, en lugar de la REST API existente en la versión anterior, permitiendo su utilización sin necesidad de conectarse al WiFi del servidor que carece de conexión a Internet.

La Fig. 4 muestra dos esquemas con la comunicación original y con los protocolos actuales.



Protocolos de comunicación originales



Protocolos de comunicación actuales

Fig. 4: Protocolos de comunicación

Mejoras en la documentación

Se documentaron todas las aplicaciones involucradas para facilitar su utilización para futuros trabajos y mejoras.

Se plantean los siguientes objetivos I+D a futuro:

- Diseñar una nueva versión del dispositivo mejorando la PCB y con nuevas funcionalidades como por ejemplo la incorporación de sonidos.
- Crear un nuevo proceso de fabricación rápido y low-cost.
- Incorporar una amplia gama de formas y colores RGB en la aplicación móvil.
- Integrar la aplicación móvil con el sistema de gestión de historias deportivas desarrollado en el LINTI (MACHI).
- Ofrecer acceso a los dispositivos para su uso en el entrenamiento en clubes y centros deportivos.
- Volver a realizar pruebas de campo en ambientes controlados.

4. Formación de Recursos Humanos

El equipo de trabajo de la línea de I+D+i presentada en este artículo se encuentra formado por docentes investigadores categorizados del LINTI y alumnos avanzados de Ingeniería en Computación perteneciente a Facultad de Informática y a la Facultad de Ingeniería.

En relación a las tesinas de grado vinculadas con IoT, se está dirigiendo a dos tesistas. Se han finalizado tres Prácticas Profesional Supervisadas (PPS) que trabajaron sobre Tremün+ y se encuentran cuatro en progreso, relacionadas con temas de IoT.

Asimismo, este año se está ejecutando el proyecto Tecnologías aplicadas al Deporte: dispositivos de bajo costo y análisis de datos, financiado por la Facultad de Informática de la UNLP.

5. Referencias

Alexandre D, da Silva CD, Hill-Haas S, et al. *Heart rate monitoring in soccer: interest and limits during competitive match play and training, practical application*. J Strength Cond Res. 2012;26(10): 2890–2906. doi:10.1519/JSC.0b013e3182429ac7

Fava, L., Vilches, D., Romero Dapozzo, R., Pagano, M., *Tecnología aplicada al deporte de alto rendimiento*, XX

Workshop de Investigadores en Ciencias de la Computación, Corrientes, Argentina, 2018.

Lamberti, C., *Brain Training Enhancing Athletic Performance*, SMARTfit Multisensory Fitness. https://www.ibm.com/developerworks/ssa/library/iot-mqtt-why-good-for-iot/index.html, Inc., Febrero 2018.

Suwardi Ansyah, A.S *et al*., *MQTT Broker Performance Comparison between AWS, Microsoft Azure and Google Cloud Platform*, 2023 International Conference on Recent Trends in Electronics and Communication (ICRTEC), doi: 10.1109/ICRTEC56977.2023.10111870.

Walton C.C.; Keegan R.J.; Martin M.; Hallock H. *The Potential Role for Cognitive Training in Sport: More Research Needed*. En: Frontiers in Psychology (2018).

Un marco de Gobernanza de Datos para la Transformación Digital en la provincia de Río Negro

Mauro Cambarieri, Alejandra Viadana, Héctor Ruiz, Nicolás Martin, Vanina Nervi, Formichella Jorge, Sofía Rached, Bruno Gorosito, Fernando González

Universidad Nacional de Rio Negro. Sede Atlántica
Laboratorio de Informática Aplicada
{mcambarieri, caviadana, hruiz, namartin, vmnervi,jformichella,msrachedgalera,bjgorosito}@unrn.edu.ar
fernando.g.gonzalez@hotmail.com

RESUMEN

Este proyecto tiene como propósito establecer un marco de trabajo para la gestión de datos que impulse la transformación digital en el sector público, considerando que los datos se han convertido en un activo estratégico fundamental para las instituciones gubernamentales. En particular, se busca diseñar un marco que guíe a las instituciones públicas de la provincia de Río Negro en la adopción de un modelo de gobernanza de datos, capaz de responder a las necesidades de cada sector y cumplir con los requisitos normativos y regulatorios. El resultado esperado es la implementación de un modelo de gobernanza de datos sólido y eficaz, que estandarice procesos, mejore la calidad y seguridad de los datos, y fomente la interoperabilidad. Esto permitirá la toma de decisiones basada en evidencia, para una mejor prestación de servicios, mayor transparencia, rendición de cuentas, e innovación en el sector público. El impacto en la gestión de datos será significativo, ya que se espera: 1) Definir roles y responsabilidades en la gestión de datos, incluyendo la figura del Chief Data Officer (CDO). 2) Clasificar los datos y establecer protocolos de acceso y formatos. 3) Generar estándares que faciliten la interoperabilidad. 4) Definir lineamientos para el uso ético de los datos. 5) Establecer principios que guíen la calidad de los datos en todo su ciclo de vida. Este proyecto se alinea con las mejores prácticas internacionales, propuesta por la Organización para la Cooperación y el Desarrollo Económicos (OCDE), y busca adaptarlas al contexto específico de la administración pública provincial. Además, la OCDE ha elaborado un plan de trabajo que ofrece recomendaciones a los países miembros en temas de gobernanza de datos, considerando políticas de privacidad, intercambio de datos transfronterizos y acceso a la información, entre otros aspectos. Este Marco de Gobernanza permitirá que las agencias de gobierno evolucionen hacia organizaciones orientadas por los datos, promoviendo la gestión estratégica de la información para mejorar los servicios públicos locales y provinciales, alcanzando mayor transparencia, interacción, integración, acceso, estandarización, intercambio y gestión de los datos, independientemente de la entidad que los posea, y facilitando la toma de decisiones que impacte positivamente en la calidad de vida de los ciudadanos

Palabras clave: Gobernanza de datos, Transformación digital, Gestión de Datos, Gobierno digital.

CONTEXTO

En un contexto cada vez más digitalizado, los datos se han convertido en uno de los activos más valiosos para las organizaciones, especialmente en el sector público, donde los gobiernos municipales y provinciales buscan mejorar los servicios, la calidad de vida de los ciudadanos y la protección de la información. Contar con políticas públicas basadas en datos permitirá establecer lineamientos sólidos para una adecuada gobernanza de datos, sentando las bases para una política integral en cuanto a la generación, protección, intercambio, uso, análisis y gestión de la información. Sin embargo, la falta de mejores prácticas, metodologías y normativas, junto con la limitada disponibilidad de recursos humanos capacitados, tecnológicos y financieros, dificulta que los gobiernos puedan ofrecer soluciones rápidas y efectivas ante los desafíos de la gobernanza de datos. En este escenario, los recursos tecnológicos se presentan tanto como facilitadores esenciales y herramientas de innovación para mejorar la calidad, seguridad y cumplimiento de los datos gubernamentales.

A continuación, se presentan algunos de los antecedentes que el grupo de investigación viene trabajando en el tema:

- PI 40-C-875 “Herramientas Informáticas de Dominio Específico para el Desarrollo de Servicios Digitales Innovadores para Comunidades Urbanas

y Rurales en el Marco de Ciudades y Regiones Inteligentes" desarrollado en el Laboratorio de Informática Aplicada, Sede Atlántica, Universidad Nacional del Río Negro (UNRN).

- Publicación en Electronic Journal of SADIO 2025: "Microcredentials and Web3: Exploring the Potential and Opportunities for Digital Transformation"
- CACIC 2024: "Explorando el potencial de las microcredenciales y la tecnología Blockchain para la Transformación Digital en la Educación Superior" http://rid.unrn.edu.ar/handle/20.500.12049/12336
- Workshop de Investigadores en Ciencias de la Computación 2024: "Aplicación de tecnología Blockchain para la emisión y verificación de Microcredenciales" http://rid.unrn.edu.ar/handle/20.500.12049/12335
- Workshop de Investigadores en Ciencias de la Computación 2023: "Diseñar una solución de identidad auto-gestionada para acceso a servicios de calidad con redes Blockchain multipropósito en la Universidad Nacional de Río Negro" https://rid.unrn.edu.ar/handle/20.500.12049/11011
- (JAIIO) "Adoptando el nuevo paradigma de producción de software con redes Blockchain multipropósito para el diseño de una solución de Identidad Digital Autogestionada en la Universidad Nacional de Río Negro". https://rid.unrn.edu.ar/handle/20.500.12049/11011
- (JAIIO). Implementación de la Blockchain del Padrón Federal en la Agencia De Recaudación". http://rid.unrn.edu.ar/handle/20.500.12049/9699
- Presentación del proyecto: "Explorando el potencial de las microcredenciales y la tecnología Blockchain para la Transformación Digital en la Educación Superior" para la convocatoria PICTO Patagonia Norte de la Agencia Nacional de Promoción de la Investigación, el Desarrollo Tecnológico y la Innovación (Agencia I+D+i)

Los beneficios de la creciente adopción del Gobierno Digital incluyen: una mayor accesibilidad y disponibilidad de los servicios públicos, agilización de trámites administrativos, fortalecimiento de la participación ciudadana, y un aumento en la transparencia y eficiencia del gobierno, además de promover la innovación en la administración pública (United Nations, 2019). Las iniciativas emergentes y tendencias actuales abarcan la adopción de servicios en línea, datos abiertos, digitalización de trámites y procesos administrativos, trazabilidad y transparencia de la información mediante Blockchain, y el uso de Inteligencia Artificial con agentes conversacionales, entre otros. Todas estas prácticas utilizan los datos como un activo estratégico para el progreso de cualquier institución pública. El proyecto "Un Modelo de Gobernanza de Datos para la Transformación Digital" tiene como objetivo ayudar a las organizaciones a gestionar y utilizar sus datos de manera segura, interoperable, eficiente y transparente, logrando soluciones más efectivas y adaptadas a las necesidades y requisitos del sector público en particular.

Este proyecto busca crear un marco de trabajo para la gestión de datos, que incluya la definición de roles y responsabilidades, la estandarización de procesos, y el desarrollo de capacitación y herramientas/metodologías para asegurar la calidad y seguridad de la información. Entre los beneficios esperados se encuentran una mayor eficiencia en la gestión pública, servicios más personalizados y adaptados a las necesidades regionales, mayor transparencia en la toma de decisiones, y un aumento en la participación ciudadana.

1. INTRODUCCIÓN

La gobernanza de datos se ha convertido en un componente clave de la política pública en relación con la Transformación digital de las organizaciones.

Para llevar a cabo un gobierno de datos, estos deben sustentarse en marcos de trabajos solidos que garanticen el acceso, la integración, el procesamiento, el almacenamiento de manera ética y segura, además de establecer políticas y normativas claras para definir estándares para la gestión y uso. A partir de estos marcos sólidos de gobernanza de datos, que maximizan el potencial de estos, permiten formular políticas para mejorar la toma de decisiones con base empírica, la entrega de servicios digitales y, en última instancia, la calidad de vida de los ciudadanos. [1].

La importancia radica en la necesidad de establecer reglas, políticas, principios y normativas que administren el ciclo de vida de los datos, determinando los roles y responsabilidades de los actores involucrados en su uso y reutilización [2].

En los últimos años se están observando esfuerzos por valorizar los datos de manera integral, como un insumo crítico para la toma de decisiones donde se distinguen principalmente iniciativas para generar lineamientos básicos que permitan la interoperabilidad y la transferencia de datos entre países, asegurando también su seguridad.

A nivel internacional, la discusión sobre la gobernanza de datos se encuentra en un estado incipiente. Si bien hay incremento de los debates y diversas iniciativas regionales y nacionales, sus alcances aún se limitan a proyectos específicos [1].

La Unión Europea, propuso en el año 2020 una regulación para la gobernanza de datos que busca incrementar la transferencia de datos entre los países miembros, permitiendo que la sociedad y las instituciones públicas y privada operen en su conjunto

y aprovechen el potencial de los mismos [3]. Dentro de los esfuerzos internacionales, también se destaca el marco conceptual para la gobernanza de datos introducido por la OCDE en 2019 [4]. Este modelo estudió los esfuerzos para una mejor gobernanza de los datos y capturó los elementos necesarios en tres niveles: estratégico, táctico y de implementación.

En América Latina, la agenda de gobernanza de datos ha ganado terreno en los últimos años, pero aún enfrenta desafíos importantes. Diversos países de la región se destacan en la apertura de datos, pero pocos han logrado generar una cultura de datos y contar con un programa integral de gobernanza [1].

En los últimos años, impulsados por iniciativas multilaterales y regionales como la Unión Europea, numerosos países de la OCDE [5] han avanzado hacia la adopción de modelos de gobernanza de datos integrales a nivel nacional. Un número creciente de gobiernos está incorporando estrategias nacionales y marcos conceptuales que buscan la integración de las diferentes agencias del sector público, con el objetivo de promover políticas públicas sólidas y maximizar el aprovechamiento estratégico de los datos.

Hoy en día se observa una multiplicación de iniciativas orientadas a la gobernanza de datos, reflejo de la creciente importancia de este campo. Sin embargo, el avance de los proyectos varía significativamente entre países, influenciado por sus contextos locales y capacidades institucionales. Algunos ejemplos destacados a continuación ilustran enfoques innovadores y estrategias integrales para optimizar la gestión y el uso de datos en el sector público. Estas iniciativas resaltan la necesidad de adaptar los modelos de gobernanza a las realidades específicas de cada nación o región, garantizando su eficacia y sostenibilidad.

Diversos avances permiten alcanzar la transformación digital, la gobernanza de datos y el desarrollo de servicios digitales innovadores. En este sentido, la gobernanza de datos se ha consolidado como un pilar estratégico en la transformación digital, abordando la necesidad de gestionar de manera efectiva el ciclo de vida de los datos, desde su generación hasta su eliminación, para garantizar su calidad, seguridad e interoperabilidad. A través de diversos marcos conceptuales y estudios de caso, se han identificado buenas prácticas, modelos exitosos y desafíos persistentes en este campo. Algunos modelos y avances relevantes incluyen: Reino Unido[1]: Cuenta con una estrategia nacional de datos desde 2020, basada en cuatro pilares: las bases, las habilidades, la disponibilidad y el uso responsable de los datos; Dinamarca[2]: desarrolló la Estrategia de Crecimiento Digital 2025, que incluye el programa "Datos como motor para el Crecimiento", cuyo objetivo es facilitar datos públicos al sector privado; Noruega[3]: la Agencia para la Gestión Pública y Gobierno Electrónico creó un modelo de gobernanza de la información que posiciona la gestión de datos en el centro de la transformación digital del sector público; Nueva Zelanda[4]: su nuevo marco de gobernanza adopta el "enfoque del ciclo de vida completo de los datos", que anima a los organismos públicos y empleados a pensar estratégicamente sobre el gobierno, la gestión, la calidad y la responsabilidad de sus datos. En cuanto a las experiencias en Ciudades: Londres[5]: publicó en 2022 la estrategia "Nacional de Datos", con un plan de trabajo para promover el uso de los datos y la infraestructura digital; Dubai[6]: implementó la Ley Dubai Data (2015) para definir mecanismos que faciliten la compartición de datos públicos de manera fácil y segura; Buenos Aires[7]: desarrolló el "Kit de Gobernanza de Datos", que establece una hoja de ruta para convertir a la ciudad en una organización que maximice el potencial de los datos; Ciudad de México[8]: publicó la "Política de Gestión de Datos", que establece los lineamientos que debe seguir la administración pública al gestionar los datos; Bogotá[9]: se encuentra trabajando en una guía con lineamientos para la gobernanza de datos.

Las transformaciones en gobierno tendientes a proveer ecosistemas que permiten la seguridad, transparencia y gestión de los datos para acceder a recursos e innovaciones a través de la colaboración con otros actores, en relación a los marcos internaciones: la OCDE (2019) un marco conceptual basado en tres niveles (estratégico, táctico y operativo), destacando herramientas como liderazgo, infraestructura y marcos regulatorios para la gobernanza de datos[6]; por otro lado, la Unión Europea propuso “Data Governance Act” [3], que fomenta la transferencia segura e interoperable de datos entre estados miembros, en el caso de Estonia, una innovación tecnológica que permitió el desarrollo de la plataforma X-Road [7] que ejemplifica la interoperabilidad transfronteriza, asegurando el intercambio y control de los datos entre organismos y países.

[1] https://www.itgovernance.co.uk/data-governance
[2] https://digital-skills-jobs.europa.eu/en/actions/national-initiatives/national-strategies/denmark-digital-growth-strategy-2025
[3] https://www.regjeringen.no/en/dokumenter/meld.-st.-22-20202021/id2841118/?ch=1#:~:text=The%20Government%20will%20facilitate%20a%20responsible%20data%20economy%20in%20Norway,Norwegian%20and%20international%20businesses%20facilitated.
[4] https://data.govt.nz/leadership/strategy-and-roadmap
[5] https://www.gov.uk/guidance/national-data-strategy
[6] https://www.digitaldubai.ae/data/regulations
[7] https://buenosaires.gob.ar/sites/default/files/2023-09/Kit%20de%20Gobernanza%20de%20Datos_1.pdf
[8] https://politicadedatos.cdmx.gob.mx/sobre/
[9] https://www.mintic.gov.co/portal/inicio/Sala-de-prensa/Noticias/198952:MinTIC-expide-el-Plan-Nacional-de-Infraestructura-de-Datos-que-impulsara-la-transformacion-digital-del-Estado

Existen trabajos relacionados en la literatura que definen la Gobernanza de Datos (GD) como un marco que organiza derechos y responsabilidades en la toma de decisiones sobre la gestión y uso de los mismos. Según Weber [8] y Khatri y Brown [9], esta gobernanza se centra en establecer estructuras y lineamientos para regular estas decisiones y su implementación. Por otro lado, Begg y Caira [10] amplían el concepto al incluir procesos relacionados con la administración, calidad y seguridad de los datos, lo que refleja la naturaleza multidimensional del término. El concepto GD se asocia al desarrollo de un conjunto de reglas que abarcan todo el ciclo de vida de los datos, desde su recopilación, depuración, análisis e interpretación, hasta la elaboración de informes, visualización, mejora continua y operaciones.

Siguiendo esta perspectiva, el autor Otto [11] describe la gobernanza de datos como el conjunto de procesos organizativos destinados a especificar la asignación de decisiones y responsabilidades relacionadas con los datos, alineándolas con los objetivos estratégicos de la organización para fomentar un manejo eficiente y responsable de este activo.

En términos generales, existe consenso en torno a tres aspectos claves: (1) Los datos como un activo estratégico que requiere una gestión adecuada. (2) Delimitar responsabilidades en la toma de decisiones y las tareas operativas asociadas. (3) Establecer normativas que garanticen la calidad de los datos y promuevan su uso ético y eficiente [11].

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Este análisis se centrará en la transformación digital en el sector público, explorando el diseño y desarrollo continuo de servicios digitales innovadores para la gestión y gobernanza de datos. La revisión incluirá un análisis de publicaciones científicas, estudios de caso implementados por gobiernos en los diferentes niveles, y recomendaciones de políticas emitidas por los gobiernos nacionales, provinciales y locales, así como por organismos internacionales.

La hipótesis central del proyecto sostiene que la implementación de un modelo de gobernanza de datos sostenible, fundamentado en mejores prácticas y el fortalecimiento de capacidades, mejorará significativamente la calidad, seguridad y cumplimiento normativo de los datos, fomentando una mayor transparencia y confianza pública. Este marco abordará tanto la definición de roles y responsabilidades en la gestión de datos como la estandarización de procesos y la creación de lineamientos éticos, lo cual impulsará la interoperabilidad y permitirá la rápida adopción de tecnologías emergentes.

Objetivo General: Desarrollar un marco de gobernanza de datos para la provincia de Río Negro que promueva la calidad, seguridad, intercambio y cumplimiento normativo de los datos, mediante la adopción de mejores prácticas, el fortalecimiento de capacidades de los agentes públicos y el aumento de la confianza en la gestión de la información. En este contexto, el proyecto busca la definición de mejores prácticas, para que los gobiernos municipales y provinciales puedan innovar de manera ágil en la prestación de sus servicios, aprovechando los datos como un activo estratégico para el desarrollo regional. En base a esta premisa, se definen los siguientes objetivos específicos.

- O1: Definir la estrategia de gobernanza de datos, incluyendo la misión, principios rectores, ejes estratégicos y líneas de acción a implementar.
- O2: Diseñar y coordinar la implementación de la estrategia de gobernanza de datos en la provincia.
- O3: Establecer estándares e instrumentos para la gestión de datos, como estándares de interoperabilidad, un marco de ética, y herramientas para la protección de los datos.
- O4: Investigar el estado del arte en nuevas tecnologías para la gestión de datos y la entrega de servicios digitales integrados basados en datos compartidos.
- O5: Desarrollar competencias y habilidades en gobernanza de datos fortaleciendo el talento humano para responder a los desafíos de la gestión de datos.
- O6: Difundir conocimiento sobre la gestión de datos mediante documentos, estándares, buenas prácticas y servicios de asistencia técnica.
- O7: Desarrollar herramientas, metodologías y mejores prácticas que optimicen el diseño y uso de los datos, promoviendo la entrega ágil de servicios innovadores que respondan a las necesidades de la provincia de Río Negro.
- O8: Impulsar comunidades de prácticas que refuercen la adopción y desarrollo de la gobernanza de datos en el gobierno local y provincial.
- O9: Proponer un marco de trabajo en gestión de datos para la toma de decisiones basadas en evidencia, que facilite una transformación digital rápida en los ámbitos municipal y provincial. Este marco se desarrollará a partir de la comparación del estado del arte y las herramientas, metodologías y mejores prácticas mencionadas en los objetivos anteriores.

3. RESULTADOS OBTENIDOS/ ESPERADOS

- Desarrollo de material educativo para cursos de grado y posgrado sobre gobernanza de datos.
- Definición clara de roles y responsabilidades en la gestión de datos, incluyendo la figura del Chief Data Officer.

- Establecimiento de protocolos para el acceso, clasificación y formatos de datos.
- Creación de estándares que faciliten la interoperabilidad entre organismos.
- Elaboración de lineamientos para el uso ético de los datos y buenas prácticas en el sector público.
- Identificación y adopción de herramientas, enfoques y soluciones innovadoras para la gestión de datos.
- Desarrollo de metodologías que optimicen la administración y calidad de los datos.
- Capacitación de agentes públicos y funcionarios en principios y prácticas de gobernanza de datos.
- Definición de políticas y lineamientos que regulen la gestión de datos en cumplimiento con normativas vigentes.
- Implementación de un Portal de Datos Abiertos que facilite la publicación y acceso a información pública relevante.
- Definición, diseño y estrategias para la creación de una Oficina de Datos que fortalezca la transformación digital en las instituciones públicas

4. FORMACIÓN DE RECURSOS HUMANOS

El grupo de trabajo se encuentra formado por cuatro investigadores formados, dos investigadores en formación y dos alumnos avanzados de la carrera Licenciatura en Sistemas. En su marco se desarrollarán prácticas profesionales supervisadas, becas de formación práctica y se producirán dos trabajos finales de carrera de grado y la elaboración del plan de trabajo de la carrera de Doctorado – Sede Atlántica.

5. BIBLIOGRAFÍA

[1]. S. Cabello, “Análisis de los modelos de gobernanza de datos en el sector público: una mirada desde Bogotá, Buenos Aires, Ciudad de México y São Paulo”, Documentos de Proyectos (LC/TS.2023/71), Santiago, Comisión Económica para América Latina y el Caribe (CEPAL), 2023.

[2]. Modelo de Gobernanza de la Infraestructura de Datos para el Estado Colombiano
Acción A1: Plan Nacional de Infraestructura de Datos (PNID) – 2022. Elaborado por: PolicyLab

[3]. Ley Europea de Gobernanza de Datos. Accedido 1/11/2024 disponible en https://digital-strategy.ec.europa.eu/es/policies/data-governance-act

[4] OECD (2019), The Path to Becoming a Data-Driven Public Sector, OECD Digital Government Studies, OECD Publishing, Paris, https://doi.org/10.1787/059814a7-en.

[5]. OECD (2019), "Data governance in the public sector", in The Path to Becoming a Data-Driven Public Sector, OECD Publishing, Paris, https://doi.org/10.1787/9cada708-en.

[6]. Datos y Gobierno Digital. Banco Interamericano de Desarrollo. Disponible en https://www.iadb.org/es/quienes-somos/topicos/modernizacion-del-estado/datos-y-gobierno-digital

[7]. X-Road, Disponible en https://e-estonia.com/solutions/interoperability-services/x-road/.

[8] Weber, K., Otto, B. y Österle, H. (2009). One Size Does Not Fit All. A Contingency Approach to Data Governance. Journal of Data and Information Quality, 1(1), 1-26. https://doi.org/10.1145/1515693.1515696

[9] Khatri, V. y Brown, C. V. (2010). Designing Data Governance. Communications of the ACM, 53(1), 148-152. https://doi.org/10.1145/1629175.1629210

[10] Begg, C. y Caira, T. (2012). Exploring the SME Quandary: Data Governance in Practise in the Small to Medium-Sized Enterprise Sector. Electronic Journal of Information Systems Evaluation, 15(1), 3-13.

[11] Otto, B. (2011b). Organizing Data Governance: Findings from the Telecommunications Industry and Consequences for Large Service Providers. Communications of the Association for Information Systems, 29, 45-66 (artículo 3). https://doi.org/10.17705/1cais.02903

· PSSTR ·

Procesamiento de Señales y Sistemas de Tiempo Real

APLICACIÓN DE PLANIFICACIÓN DE TIEMPO REAL HETEROGÉNEA EN SISTEMAS EMBEBIDOS, IOT Y ROBÓTICA

Francisco Ezequiel Páez[1], José M. Urriza[1], Javier Orozco[2,4], Ricardo Cayssials[3], Fabricio Pafumi[1], Jonathan Cavia[1], Matias Castiglione[1]

[1] Depto. de Informática, Facultad de Ingeniería, Universidad Nacional de la Patagonia San Juan Bosco, Sede Puerto Madryn

[2] Depto. de Ingeniería Eléctrica y de Computadoras, Universidad Nacional del Sur

[3] Universidad Tecnológica Nacional, Facultad Regional Bahía Blanca.

[4] CONICET

fep@ing.unp.edu.ar, josemurriza@gmail.com, jorozco@uns.edu.ar, rcayssials@frbb.utn.edu.ar

RESUMEN

El objetivo de este proyecto es evaluar la aplicabilidad, complejidad y perfomance de algoritmos de planificación de tiempo real heterogénos en sistemas embebidos, internet de las cosas y robótica. Gran parte de los algoritmos de planificación disponibles en la literatura son evaluados mediante simulaciones o pruebas sintéticas en plataformas de desarrollo. Por lo tanto, contar con un análisis de su implementación y comportamiento en aplicaciones concretas es de gran útilidad para evaluar la aptitud de un algoritmo. Mediante la implementación de diversos algoritmos en, por ejemplo, un brazo robot, se busca evaluar su impacto en el funcionamiento global del sistema. Se planea desarrollar un banco de pruebas con diversas aplicaciones que permitan evaluar y comparar algoritmos de planificación actuales y futuros.

Palabras clave: *Planificación*, *STR*, *Sistemas Embebidos*.

CONTEXTO

Esta línea de investigación aplicada se enmarca en el proyecto de investigación *Aplicación de Planificación de Tiempo Real Heterogénea en Sistemas Embebidos, IoT y Robótica* (80020170100007UP, Res. CDFI 198 / 2024, Red. CDFI 274 / 2024).

1. INTRODUCCIÓN

La planificación de Sistemas de Tiempo Real (STR) es ampliamente estudiada desde el trabajo seminal de Liu y Layland [1], como lo demuestran los periódicos trabajos de revisión del estado del arte ([2], [3], [4], [5], [6]). En particular la planificación de *STR heterogéneos*, que reunen tareas críticas y no críticas, ha cobrado relevancia en las últimas décadas ([7], [8], [9], [10]). Entre las técnicas de planificación más utilizadas se encuentran servidores [11] y prioridad mixtas [12]. Otras propuestas son la administración del tiempo ocioso mediante Slack Stealing ([13], [14], [15], [16]). Cuando el conjuntos de tareas es dinámico se requiere además evaluar la planificabilidad del sistema en tiempo de ejecución, para evitar comprometer los requerimientos temporales ([17], [18], [19], [20], [21]). Los desafíos de implementación han sido tratados en diversos trabajos, con especial atención en la integración de los conceptos de *STR* con otras disciplinas como control [22] o con los requerimientos del software de base como Sistemas Operativos de Tiempo Real (*SOTR*) ([23], [24], [25]). Otros trabajos y

proyectos se centran en los desafíos de implementación y el desarrollo de bancos de pruebas ([26], [27], [28], [29], [30], [31], [32]). Este proyecto en particular busca sumar un banco de prueba orientado a la investigación que incentive la experimentación por parte de alumnos de grado avanzados.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Los *STR* se caracterizan por contar con *vencimientos*, instantes antes de los cuales las tareas deben generar el resultado esperado. Su cumpliento requiere un cuidadoso diseño e implementación del sistema, especialmente cuando una falla puede poner en riesgo la vida humana o causar graves daños económicos o ambientales. Sin embargo, persiste aún una brecha entre el diseño y la implementación de un *STR*, especialmente cuando la concurrencia e interacción de tareas y eventos no es trivial. Es común que terminos como "ejecución rápida" o "en línea" sean utilizados como sinónimos de *STR*, conceptos que están lejos de ser adecuados. En la actualidad, la masificación de los dispositivos móviles y embebidos, con tecnologías de comunicación que permiten una conectividad constante (el "Internet de las Cosas" o IoT por sus siglas en inglés) remarca la importancia de la inclusión de contenidos relacionados con *STR* en carreras de grado en computación, informática e ingeniería. Mas aún, muchos sistemas combinan actualmente tareas de tiempo real críticas con otras tareas no-críticas, lo que incrementa la dificultad de una correcta planificación. Este proyecto tiene como objetivo el desarrollo de aplicaciónes prácticas de diversas técnicas de planificación de tiempo real heterogéneas, para estudiar y evaluar su complejidad (como se integra en un sistema SOTR, beneficios y/o complicaciones que aportan al ser utilizados, etc). Estas evaluaciones se sistematizaran en un banco de pruebas de aplicaciones prácticas, para complementar resultados obtenidos mediante técnicas como la simulación. Este tipo de aplicación práctica es además sumamente atractivo para estudiantes, lo que permite consolidar su formación. Se hará énfasis en que sirvan como marco para la realización de trabajos o proyectos finales de carrera.

Al ser un proyecto orientado principalmente a la investigación aplicada, la metodología de trabajo consitirá fundamentalmente en la revisión del estado del arte de la disciplina en cuanto a planificadores de tiempo real heterogeneos, plataformas de desarrollo y sistemas operativos de tiempo real. Esta actividad se centrará tanto en resultados academicos como en desarrollos de la industria. Aunque esta revisión se planea continua, tendrá énfasis en la primera mitad del proyecto. En paralelo, se diseñaran los ejemplos prácticos de implementación que constituirán el banco de pruebas, en donde se evaluará lo relevado en la literatura. Los resultados de estas evaluaciones y los posibles aportes y mejoras que se realicen serán publicados oportunamente en congresos y/o revistas, según amerite su impacto.

Los integrantes del proyecto poseen producción en investigación en el área de STR. Han trabajado en diversas técnicas de evaluación de planificabilidad para su aplicación en línea ([21], [33], [34]). Además, han presentando mejoras sustanciales a los algoritmos de Slack Stealing para la administración del tiempo ocioso, fundamental para un planificador heterogéneo [16]. También han implementado planificadores heterogéneos en SOTR como FreeRTOS, utilizando Slack Stealing ([35], [36]).

3. RESULTADOS ESPERADOS

El objetivo general del proyecto facilitar la evaluación práctica de algoritmos de planificacion de tiempo real heterogéneos.

Los objetivos particulares son:

- Generar un banco de pruebas para evaluar algoritmos de planificación.
- Documentar y evaluar la implementación de algoritmos de planificación en aplicaciones concretas.
- Plantear mejoras y recomendaciones para su implementación.
- Fomentar la participación de estudiantes, consolidando su formación en el área y generando trabajos finales de carrera.

Las actividades a desarrollar son:

- Estudio de modelos y técnicas propuestos de planificación heterogénea en STR.
- Estudio de plataformas de desarrollo y SOTR más difundidos, para su uso en el banco de pruebas
- Definición y elaboración de un banco de pruebas para evaluar algoritmos de planificación.
- Implementación y evaluación de los métodos y técnicas de planificación estudiadas.
- Formular mejoras y/o nuevos métodos en base a los resultados obtenidos en las pruebas.
- Publicar los resultados parciales y finales del proyecto que ameriten su comunicación.

4. FORMACIÓN DE RECURSOS HUMANOS

En el marco de este proyecto participan tres alumnos de Licenciatura en Informática de la Sede Puerto Madryn, estando planeado el desarrollo de la tesina de grado de uno de estos alumnos en el marco del proyecto.

Los docentes que participan del proyecto forman parte del Grupo de Investigación sobre Sistemas de Tiempo Real en la UNPSJB Sede Puerto Madryn. Este grupo mantiene vínculos con investigadores de la misma disciplina del Departamento de Ingeniería Eléctrica y de Computadoras (DIEC) de la Universidad Nacional del Sur (UNS) y de la UTN Regional Bahía Blanca. Se espera que durante el trascurso del proyecto el grupo de investigación extienda sus vinculaciones.

5. CONCLUSIONES

Este proyecto tiene como objetivo generar un banco de pruebas para la evaluación práctica de algoritmos de planificación para *STR heterogéneos*, que complemente evaluaciones mediante simulaciones. Este enfoque permitirá evaluar el comportamiento de estos algoritmos en aplicaciones concretas, particularmente en sistemas embebidos, IoT y robótica. Los resultados que se generen ayudarán a tomar mejores decisiones a la hora de seleccionar un algoritmos de planificación en sistemas específicos. Además, la implementación de este banco de pruebas servirá como herramienta educativa en la formación de estudiantes avanzados. Al combinar aspectos teóricos y prácticos, se espera que que sea una opción atractiva para el desarrollo de trabajos finales de carrera.

BIBLIOGRAFÍA

[1] C. L. Liu y J. W. Layland, «Scheduling Algorithms for Multiprogramming in a Hard-Real-Time Environment», *J. ACM*, vol. 20, n.º 1, pp. 46-61, ene. 1973, doi: 10.1145/321738.321743.

[2] L. Sha *et al.*, «Real Time Scheduling Theory: A Historical Perspective», *Real-Time Systems*, vol. 28, n.º 2/3, pp. 101-155, nov. 2004, doi: 10.1023/B:TIME.0000045315.61234.1e.

[3] S. Baruah, M. Bertogna, y G. Buttazzo, *Multiprocessor Scheduling for Real-Time Systems*. en Embedded Systems. Cham: Springer International Publishing, 2015. doi: 10.1007/978-3-319-08696-5.

[4] M. Bambagini, M. Marinoni, H. Aydin, y G. Buttazzo, «Energy-Aware Scheduling for Real-Time

Systems: A Survey», *ACM Transactions on Embedded Computing Systems*, vol. 15, n.º 1, pp. 1-34, ene. 2016, doi: 10.1145/2808231.

[5] R. I. Davis y A. Burns, «A Survey of Hard Real-time Scheduling for Multiprocessor Systems», *ACM Comput. Surv.*, vol. 43, n.º 4, p. 35:1-35:44, oct. 2011, doi: 10.1145/1978802.1978814.

[6] R. I. Davis, «A Review of Fixed Priority and EDF Scheduling for Hard Real-time Uniprocessor Systems», *SIGBED Rev.*, vol. 11, n.º 1, pp. 8-19, feb. 2014, doi: 10.1145/2597457.2597458.

[7] A. Burns y R. I. Davis, «A Survey of Research into Mixed Criticality Systems», *ACM Computing Surveys*, vol. 50, n.º 6, pp. 1-37, nov. 2017, doi: 10.1145/3131347.

[8] S. Vestal, «Preemptive Scheduling of Multi-criticality Systems with Varying Degrees of Execution Time Assurance», IEEE, dic. 2007, pp. 239-243. doi: 10.1109/RTSS.2007.47.

[9] M. S. Mollison, J. P. Erickson, J. H. Anderson, S. K. Baruah, y J. A. Scoredos, «Mixed-Criticality Real-Time Scheduling for Multicore Systems», en *2010 10th IEEE International Conference on Computer and Information Technology*, jun. 2010, pp. 1864-1871. doi: 10.1109/CIT.2010.320.

[10] S. Baruah, «Mixed-Criticality Scheduling Theory: Scope, Promise, and Limitations», *IEEE Design Test*, vol. 35, n.º 2, pp. 31-37, abr. 2018, doi: 10.1109/MDAT.2017.2766571.

[11] J. K. Strosnider, J. P. Lehoczky, y Lui Sha, «The deferrable server algorithm for enhanced aperiodic responsiveness in hard real-time environments», *IEEE Transactions on Computers*, vol. 44, n.º 1, pp. 73-91, ene. 1995, doi: 10.1109/12.368008.

[12] R. Davis y A. Wellings, «Dual priority scheduling», IEEE Comput. Soc. Press, 1995, pp. 100-109. doi: 10.1109/REAL.1995.495200.

[13] T.-S. Tia, J. W.-S. Liu, y M. Shankar, «Algorithms and optimality of scheduling soft aperiodic requests in fixed-priority preemptive systems», *Real Time Systems*, vol. 10, n.º 1, pp. 23-43, ene. 1996, doi: 10.1007/BF00357882.

[14] Thuel y Lehoczky, «Algorithms for scheduling hard aperiodic tasks in fixed-priority systems using slack stealing», IEEE Comput. Soc. Press, 1994, pp. 22-33. doi: 10.1109/REAL.1994.342733.

[15] R. I. Davis, K. W. Tindell, y A. Burns, «Scheduling slack time in fixed priority pre-emptive systems», en *1993 Proceedings Real-Time Systems Symposium*, dic. 1993, pp. 222-231. doi: 10.1109/REAL.1993.393496.

[16] J. M. Urriza, F. E. Páez, R. Cayssials, J. D. Orozco, y L. Schorb, «Low Cost Slack Stealing Method for RM/DM», *International Review on Computers and Software*, vol. 5, n.º 6, pp. 660-667, oct. 2010.

[17] M. Joseph y P. Pandya, «Finding Response Times in a Real-Time System», *Comput J*, vol. 29, n.º 5, pp. 390-395, ene. 1986, doi: 10.1093/comjnl/29.5.390.

[18] M. Sjodin y H. Hansson, «Improved response-time analysis calculations», IEEE Comput. Soc, 1998, pp. 399-408. doi: 10.1109/REAL.1998.739773.

[19] S. Baruah, M. Bertogna, y G. Buttazzo, «Response Time Analysis: The [RTA] Test», en *Multiprocessor Scheduling for Real-Time Systems*, S. Baruah, M. Bertogna, y G. Buttazzo, Eds., Cham: Springer International Publishing, 2015, pp. 155-164. doi: 10.1007/978-3-319-08696-5_17.

[20] R. I. Davis, A. Zabos, y A. Burns, «Efficient Exact Schedulability Tests for Fixed Priority Real-Time Systems», *IEEE Transactions on Computers*, vol. 57, n.º 9, pp. 1261-1276, sep. 2008, doi: 10.1109/TC.2008.66.

[21] J. M. Urriza, F. E. Páez, M. Ferrari, R. Cayssials, y J. D. Orozco, «A new RM/DM low cost schedulability test», en *2017 Eight Argentine Symposium and Conference on Embedded Systems (CASE)*, ago. 2017, pp. 1-6. doi: 10.23919/SASE-CASE.2017.8115368.

[22] A. Gambier, «Real-time control systems: a tutorial», en *2004 5th Asian Control Conference (IEEE Cat. No.04EX904)*, jul. 2004, pp. 1024-1031 Vol.2.

[23] J. A. Stankovic y R. Rajkumar, «Real-Time Operating Systems», *Real-Time Systems*, vol. 28, n.º 2, pp. 237-253, nov. 2004, doi: 10.1023/B:TIME.0000045319.20260.73.

[24] J. L. Herman, C. J. Kenna, M. S. Mollison, J. H. Anderson, y D. M. Johnson, «RTOS Support for Multicore Mixed-Criticality Systems», en *2012 IEEE 18th Real Time and Embedded Technology and Applications Symposium*, abr. 2012, pp. 197-208. doi: 10.1109/RTAS.2012.24.

[25] V. Nandana, A. Jithendran, y R. Shreelekshmi, «Survey on RTOS: Evolution, Types and Current Research», *International Journal of Computer Applications*, vol. 121, pp. 28-31, jul. 2015, doi: 10.5120/21825-5077.

[26] J. M. Calandrino, H. Leontyev, A. Block, U. C. Devi, y J. H. Anderson, «LITMUS^RT : A Testbed for Empirically Comparing Real-Time Multiprocessor Schedulers», en *2006 27th IEEE International Real-Time Systems Symposium (RTSS'06)*, dic. 2006, pp. 111-126. doi: 10.1109/RTSS.2006.27.

[27] P. C. Richardson, W. Xiang, y S. Mohammad, «Performance Analysis of a Real-Time Control Network Test Bed in a Linux-Based System With Sporadic Message Arrivals», *IEEE Transactions on Industrial Informatics*, vol. 2, n.º 4, pp. 231-241, nov. 2006, doi: 10.1109/TII.2006.885925.

[28] K.-D. Kang, P. H. Sin, J. Oh, y S. H. Son, «A Real-Time Database Testbed and Performance Evaluation», en *13th IEEE International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA 2007)*, ago. 2007, pp. 319-326. doi: 10.1109/RTCSA.2007.12.

[29] B. B. Brandenburg y J. H. Anderson, «On the Implementation of Global Real-Time Schedulers», en *2009 30th IEEE Real-Time Systems Symposium*, dic. 2009, pp. 214-224. doi: 10.1109/RTSS.2009.23.

[30] C. Adjih *et al.*, «FIT IoT-LAB: A large scale open experimental IoT testbed», en *2015 IEEE 2nd World Forum on Internet of Things (WF-IoT)*, dic. 2015, pp. 459-464. doi: 10.1109/WF-IoT.2015.7389098.

[31] N. Michael, J. Fink, y V. Kumar, «Experimental Testbed for Large Multirobot Teams», *IEEE Robotics & Automation Magazine*, vol. 15, n.º 1, pp. 53-61, mar. 2008, doi: 10.1109/M-RA.2007.914924.

[32] D. Pickem *et al.*, «The Robotarium: A remotely accessible swarm robotics research testbed», en *2017 IEEE International Conference on Robotics and Automation (ICRA)*, may 2017, pp. 1699-1706. doi: 10.1109/ICRA.2017.7989200.

[33] J. M. Urriza, L. Schorb, J. D. Orozco, y R. Cayssials, «Reduced Computational Cost in the Calculation of Worst Case Response Time for Real Time Systems», vol. 9, n.º 2, p. 10, 2009.

[34] J. M. Urriza, F. E. Paez, J. D. Orozco, y R. Casyssials, «Computational Cost Reduction for Real-Time Schedulability Tests Algorithms», *IEEE Latin America Transactions*, vol. 13, n.º 12, pp. 3714-3723, dic. 2015, doi: 10.1109/TLA.2015.7404899.

[35] F. E. Páez, J. M. Urriza, R. Cayssials, y J. D. Orozco, «FreeRTOS user mode scheduler for mixed critical systems», en *2015 Sixth Argentine Conference on Embedded Systems (CASE)*, ago. 2015, pp. 37-42. doi: 10.1109/SASE-CASE.2015.7295845.

[36] F. E. Páez, J. M. Urriza, y J. D. Orozco, «Idle Time Administration on FreeRTOS Using Slack Stealing», en *Computer Science – CACIC 2020*, P. Pesado y J. Eterovic, Eds., Cham: Springer International Publishing, 2021, pp. 275-289. doi: 10.1007/978-3-030-75836-3_19.

DEEP LEARNING APLICADO A LA IDENTIFICACIÓN DE CANTOS DE ANUNCIO DE BOANA RIOJANA (AMPHIBIA: ANURA)

Tolaba, Norberto I.[1], Boullhesen, M.[2], Vaira, Marcos [2], Akmentins, M. S. [2], Revollo Sarmiento, G. Noelia[1,2]

[1]Laboratorio de GeoTecnologías y Ciencias de las Imágenes, Facultad de Ingeniería, Universidad Nacional de Jujuy (FI-UNJU).
[2]Instituto de Ecorregiones Andinas, Consejo Nacional de Investigaciones Científicas y Técnicas, Universidad Nacional de Jujuy (INECOA-CONICET).
{ivn.tlb, martinbulen01, mauriakme, marcos.vaira, giselarevollo}@gmail.com

RESUMEN

El monitoreo de poblaciones de anuros es un aspecto fundamental para la conservación de las especies. La identificación y el análisis de cantos de anuncio de anuros permiten obtener información valiosa para evaluar su estado y las amenazas que pueden enfrentar. Las técnicas clásicas de procesamiento de estos datos son tareas complejas, que consumen tiempo, tienen un costo asociado y se necesita del experto. El aprendizaje profundo, Deep Learning, ha ganado una importancia significativa en el procesamiento de datos complejos. Integrar estas técnicas en el análisis de audios de cantos de anuros, sin duda nos proporciona un nuevo enfoque en la identificación automatizada y extracción de patrones complejos de los cantos. En este sentido, se genera conocimiento útil y la realización de procesos iterativos más ágiles y eficientes. En este contexto y en el ámbito de la bioacústica, exploramos tres técnicas de DL para la clasificación automática del canto de anuros. Se construyeron tres clasificadores aplicando los algoritmos de las redes DNN, CNN y LSTM. Los resultados son prometedores, obteniendo una exactitud global superior al 88 % en todos los modelos.

Palabras claves: Clasificación Automática, Procesamiento de señales, Deep Learning, Boana riojana.

CONTEXTO

Este trabajo se realiza en estrecha colaboración con investigadores del centro de investigación INECOA (Instituto de Ecorregiones Andinas CONICET-UNJU) y la Facultad de Ingeniería de la Universidad Nacional de Jujuy, y el LGCI (Laboratorio de Geotecnologías y Ciencias de las Imágenes) de la Facultad de Ingeniería (FI-UNJU).

1. INTRODUCCIÓN

En la actualidad, las poblaciones de anfibios especialmente el grupo de vertebrados tetrápodos son la especie más amenazados a nivel global [1]. La declinación de poblaciones de anuros conlleva implicancias importantes para el funcionamiento de muchos ecosistemas terrestres, por ejemplo, la regulación de la abundancia de insectos [2], pudiendo repercutir en el bienestar humano [3].

El desarrollo y empleo del Monitoreo Acústico Pasivo (MAP) es la técnica de recolección de sonidos ambientales sin intervención humana directa en el momento de la grabación. La misma es utilizada en diversos ambientes [4]. El monitoreo de las poblaciones de anfibios, permiten obtener información valiosa a partir de los cantos de anuncio de las especies [5]. En este contexto, determinar patrones de actividad temporal tanto diaria como estacional o interanual es esencial [6], [7]. Tradicionalmente, la identificación de estas especies se ha realizado mediante métodos manuales o

semiautomáticos, los cuales requieren el conocimiento de un experto y son propensos a errores subjetivos.

El continuo avance de la tecnología, especialmente en el área del aprendizaje profundo, Deep Learning (DL), es una potente alternativa de solución e innovación en distintas disciplinas como ciencias de la tierra, biología, agricultura, industria y medicina [8], [9], [10]. En este sentido, resulta esencial investigar y desarrollar nuevos métodos aplicando DL que permitan abordar la clasificación automática de los cantos de anuncio de anfibios. Además, la integración de técnicas de DL en la clasificación automática de los cantos de anuncio de anuros es de suma importancia en el contexto de la bioacústica. Estas técnicas posibilitan una exploración más exhaustiva de los datos al identificar automáticamente patrones complejos y diferenciarlos del ruido ambiental.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El estudio de las poblaciones de anuros tiene implicancias importantes para el funcionamiento de los ecosistemas terrestres, pudiendo repercutir en el bienestar humano. El monitoreo de los mismos, es esencial para poder caracterizarlos temporalmente y explorar así sus relaciones con variables climáticas de diferentes lugares. La forma tradicional de abordar esta tarea es tediosa, consume tiempo y se necesita del experto para reconocer los cantos de los anuros. Particularmente, la *Boana riojana (Br),* es un anuro de la familia Hylidae. Este grupo, en su mayoría, son ranas arborícolas que tienen almohadillas largas para trepar por la vegetación [11]. El objetivo de este trabajo es identificar automáticamente el canto de anuncio de *Boana riojana* aplicando técnicas de Deep Learning.

El dataset con audios de cantos de anuncio de *Br*, fue recopilado en diferentes estaciones entre el 2017 y 2019 en el Parque Nacional Calilegua, provincia de Jujuy, Argentina. El grabador fue instalado en un sitio denominado “Arroyo Tres Cruces” que pertenece al piso de Bosque Montano Inferior, ubicado en los 23°41’36.84” S; 64°52’5.04” O, en una altitud de 1125 msnm., es un arroyo permanente dominado por un bosque perennifolio. El grabador digital automatizado modelo Song Meter SM4 (Wildlife Acoustics, Inc., Concord, MA, EE. UU.) fue ubicado a 1,5 m sobre el nivel del suelo. Estuvo activo desde septiembre de 2017 a septiembre de 2019 y fue programado para registrar 3 minutos consecutivos por hora (72 min/día) en el canal MONO [11]. Las grabaciones se almacenaron en archivos digitales de formato descomprimido (.WAV) en una resolución de 16 bits con un rango de frecuencia de 16 kHz reduciendo la frecuencia máxima registrada a 8 kHz, con el fin de preservar los sonidos emitidos por el ensamble de anuros de los sitios monitoreados. Las grabaciones fueron analizadas en laboratorio utilizando el software Raven Pro © 1.5 [12].

Estos audios tienen una duración de un minuto, en el cual se identifican los cantos de la especie en estudio asociado a los otros componentes del paisaje sonoro (biofonía, goofonía y antropofonía) que lo denominaremos ruido.

En la primera etapa, se realizó un pre procesamiento de los datos que consistió en los siguientes pasos: 1) Se computó la duración del canto promedio de Br, es decir, se identificaron y registraron las duraciones de los cantos en cada audio y finalmente se promedió con el total, obteniendo un valor promedio de canto de 0.760 segundos. 2) Los audios se caracterizaron y etiquetaron en dos clases de cantos de anuncios: contiguos (C), separados (S) y solo ruido (R). 3) Se analizó la duración de los cantos S y se determinó que la mínima distancia entre ellos era de 0.4 segundos. Esto permitió, hacer subdivisiones en las clases C y S, fragmentándolas en audios de 1 segundo. Esto garantizó que al realizar los cortes automáticos del sonido en las clases, cada fragmento de 1 segundo contenga el canto de Br. La clase R

directamente se dividió en fragmentos de 1 segundo. Finalmente, el dataset quedó conformado por 2,120 fragmentos de audio de 1 segundo con presencia de Boana riojana y 2,180 con ruido.

En la segunda etapa, se procesaron los audios aplicando la técnica procesamiento de señales, denominada MFCC (Mel-Frequency Cepstral Coefficients). Esta técnica permite manipular la señal y extraer características del audio, generando así una matriz de coeficientes que capturan la forma del espectro de potencia de la señal de sonido. Para ello, se separa la señal en pequeños tramos. Luego, a cada tramo se le computa la Transformada de Fourier discreta y se obtiene la potencia espectral de la señal. A continuación, se aplica el filtro de la escala Mel al espectro obtenido en el paso anterior y se suman las energías en cada uno de ellos. Se toma el logaritmo de todas las energías de cada frecuencia Mel y finalmente se aplica la transformada de coseno discreta a estos logaritmos. Este proceso se realizó mediante la implementación del paquete de Python "Librosa" para análisis de música y audios. Finalmente, se normalizaron los coeficientes para que tengan una media cercana a cero y una desviación estándar unitaria. De este modo, se evitan las diferencias de magnitud entre los datos, mejorando la estabilidad del gradiente y acelerando la convergencia del modelo durante el entrenamiento.

En la tercera etapa, se procedió a la clasificación. Para ello, se implementaron tres modelos de DL: DNN (Deep Neural Network), CNN (Convolutional Neural Network) y LSTM (Long Short-Term Memory). El modelo DNN permite modelar relaciones complejas en los datos y es utilizado para problemas de clasificación. La red LSTM resulta ideal para capturar patrones rítmicos y secuenciales de los datos. Finalmente, la CNN está diseñada para procesar datos con una estructura de cuadrícula y es ampliamente utilizada en

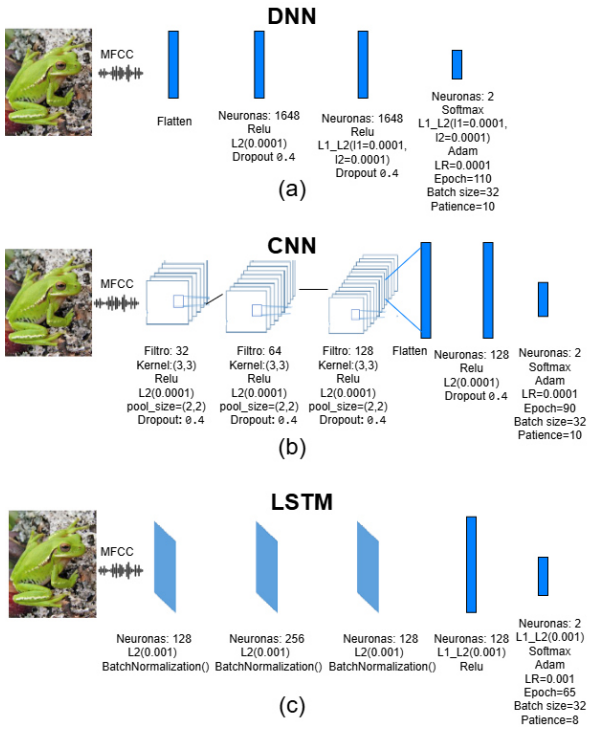


Figura 1. Arquitectura e hiperparámetros de tres redes profundas: (a) DNN. (b) CNN. (c) LSTM.

tareas de clasificación. El dataset obtenido estuvo balanceado y se dividió en un 70% para entrenamiento, y un 30% para prueba. La arquitectura e hiperparámetros de los tres modelos se especifican en la Figura 1.

En la cuarta etapa se analizó el desempeño de la clasificación. Para ello se computaron las métricas de evaluación estándar: a) Exactitud (Ag) que mide la proporción de predicciones correctas en relación con todas las predicciones realizadas. b) Precisión (precision) que evalúa cuántos de los elementos clasificados como positivos realmente son correctos. c) Recall (sensibilidad), mide qué tan bien el modelo encuentra todos los casos positivos dentro del conjunto de datos. d) F1-score. e) La matriz de confusión es el enfoque más utilizado en la evaluación de la precisión de una clasificación [13]. En la diagonal principal se reportan los casos correctamente clasificados y en la opuesta se detallan los errores de la predicción.

3. RESULTADOS OBTENIDOS/ ESPERADOS

De un dataset de 4300 audios de cantos de anuros, se desarrollaron tres modelos de redes profundas que clasifican si en una grabación está presente el canto del anuncio de *Boana riojana* (clase B) o no (clase Otro). Las matrices de confusión y los parámetros de precisión se detallan en la Tabla 1. Los resultados obtenidos en los tres modelos presentaron una exactitud global (Ag) superior al 88 %, con precisiones entre 0.89 y 0.99, prediciendo con una alta probabilidad de aciertos. El valor de Recall es superior a 0.87, indicando que todos los modelos identifican bastante bien el canto de *Boana riojana*. En todos los casos, se observa que la clase (B) se identifica correctamente. Sin embargo, se observa que la mayor tendencia a la confusión se genera en el modelo LSTM (Tabla 1(f)). En general, los resultados obtenidos en los tres modelos de clasificación, presentaron altas exactitudes y precisiones al identificar el canto de anuncio de *Br* del resto de los componentes del paisaje.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo está formado por investigadores de INECOA, un becario posdoctoral CONICET y un alumno de grado que está realizando la Práctica Profesional Supervisada (PPS) en el INECOA. Este trabajo constituye el resultado de la PPS del alumno de Ingeniería Informática Norberto Iván Tolaba y un avance en la línea de investigación abordada.

Tabla 1: Matriz de confusión y Parámetros de precisión: (a) y (b): DNN. (c) y (d): CNN. (e) y (f): LSTM

(a)

		Predicción		
		B	**O**	**Tot**
Act.	**B**	286	12	298
	O	19	328	347
	Tot	283	362	645

(b)

Parm.	**Result**
Ag [%]	95
Precisión	0.95
Recall	0.95
F1-score	0.95

(c)

		Predicción		
		B	**O**	**Tot**
Act.	**B**	281	17	298
	O	2	345	347
	Tot	283	362	645

(d)

Parm.	**Result**
Ag [%]	97
Precisión	0.97
Recall	0.96
F1-score	0.97

(e)

		Predicción		
		B	**O**	**Tot**
Act.	**B**	260	38	298
	O	29	318	347
	Tot	283	362	645

(f)

Parm.	**Result**
Ag [%]	88
Precisión	0.89
Recall	0.89
F1-score	0.89

5. BIBLIOGRAFÍA

[1] Luedtke, J. A., Chanson, J., Neam, K., Hobin, L., Maciel, A. O., Catenazzi, A., Borzée, A., Hamidy, A., Aowphol, A. & Stuart, S. N. (2023). Ongoing declines for the world's amphibians in the face of emerging threats. Nature 622 (7982), 308-314.

[2] Wells K.D. y Schwartz J.J. (2007). The Behavioral Ecology of Anuran Communication. En: P.M. Narins, A.S. Feng,

R.R. Fay, A.N. Popper (Eds.), Hearing and Sound Communication in Amphibians (pp. 44-86). Springer Handbook of Auditory Research, vol. 28. Springer, New York. EEUU.

[3] Gardner, T. (2001). Declining amphibian populations: a global phenomenon in conservation biology. Animal biodiversity and conservation, 24(2), 25-44.

[4] Szymański, P., Olszowiak, K., Wheeldon, A., Budka, M., y Osiejuk, T.S. (2021). Passive acoustic monitoring gives new insight into year-round duetting behaviour of a tropical song bird. Ecological Indicators, 122: 107271.

[5] Boullhesen, M., Vaira, M., Barquez, R. M., &Akmentins, M. S. (2021). Evaluating the efficacy of visual encounter and automated acoustic survey methods in anuran assemblages of the Yungas Andean forests of Argentina. *EcologicalIndicators*, *127*, 107750.

[6] Guerra, V., de Queiroz Costa, N., Llusia, D., Márquez, R. y Bastos, R.P. (2020). Nightly patterns of calling activity in anuran assemblages of the Cerrado, Brazil. Community Ecology, 21(1): 33-42.

[7] Boullhesen, M., Vaira, M., Barquez, R. M., &Akmentins, M. S. (2023). Patterns of acoustic phenology in ananuran assemblage of the Yungas Andean forests of Argentina. *Acta Herpetologica*, *18*(1), 23-36.

[8] Santecchia, G. S., Revollo Sarmiento, G. N., Genchi, S. A., Vitale, A. J., &Delrieux, C. A. (2023). Assessmentof Landsat-8 and Sentinel-2 Water Indices: A Case Study in the South west of the Buenos Aires Province (Argentina). *JournalofImaging*, *9*(9), 186.

[9] Revollo Sarmiento, G. N., Sarmiento, N. V. R., Delrieux, C. A., & Perillo, G. M. (2020). Morphological characterization of ponds and tidal courses in coastal wetlands using Google Earth imagery. *Estuarine, Coastal and ShelfScience*, *246*, 107041.

[10] Revollo, N. V., Sarmiento, G. N. R., Delrieux, C., Herrera, M., & González-José, R. (2021). Supervised machine learning classification of human sperm head based on morphological features. In *Trends and AdvancementsofImage Processing and ItsApplications* (pp. 177-191). Cham: Springer International Publishing.

[11] Márquez, R., Llusia, D. y Beltrán, J.F. (2014). Aplicación de la bioacústica al seguimiento de anfibios. Boletín de la Asociación Herpetológica de España, 25(2): 52-58.

[12] Center for Conservation Bioacoustics. (2014). Raven Pro: Interactive Sound Analysis Software, Version 1.5. Computer software. The Cornell Lab of Ornithology, Ithaca, New York. Disponible en: http://www.birds.cornell.edu/raven.

[13] R. G. Congalton and K. Green, Assessing the accuracy of remotely sensed data: principles and practices. Boca Raton: CRC press, 2.

Desarrollo de un Sistema Robótico Autónomo para la competición de Fútbol Robótico

Balich Néstor Adrian, Balich Franco Adrian
Asuar Juan Fernando, Badano Maximiliano Exequiel

Laboratorio de robótica física e inteligencia artificial
Universidad Abierta Interamericana
Facultad de Tecnología Informática

{nestor.balich, francoadrian.balich}@uai.edu.ar, @uai.edu.ar
{juanfernando.asuar, maximiliano.badano}@alumnos.uai.edu.ar

RESUMEN

Este trabajo presenta el desarrollo de un robot autónomo motorizado con visión artificial mediante inteligencia artificial, diseñado para simular una actividad de fútbol robótico.

Utilizando un sistema de visión artificial basado en TensorFlow Lite [1] y una aplicación desarrollada en Android [2], el robot es capaz de detectar y localizar una pelota, orientarse hacia un arco, y empujar la pelota hacia este. Este sistema tiene el potencial de impactar en la investigación robótica al ofrecer soluciones accesibles para la detección de objetos en tiempo real. Además, sus aplicaciones futuras podrían extenderse a otras áreas como la automatización industrial, la exploración en entornos desafiantes y el desarrollo de vehículos autónomos para tareas específicas, se desarrolló para movile con el fin de reducir costos y propiciar la difusión de la investigación al utilizarlo como controlador principal.

A lo largo del proyecto, se exploraron estrategias de entrenamiento de modelos de inteligencia artificial, integración de hardware, y optimización de algoritmos para el control del movimiento. Las pruebas realizadas demostraron un desempeño funcional del sistema en entornos controlados.

Palabras clave: Robótica, reconocimiento de imágenes, TensorFlow Lite, Android, control autónomo, mobile

CONTEXTO

Este proyecto se desarrolló en la cátedra de Robótica de la carrera de Ingeniería en Sistemas Informáticos de la Universidad Abierta Interamericana (UAI) dentro del laboratorio de robótica física e inteligencia artificial (LRFIA), como parte de una propuesta de investigación orientada a fomentar el aprendizaje práctico y a sentar las bases para la participación en una nueva categoría de competición de fútbol robot en la roboliga argentina, que dará inicio en 2025.

El desarrollo del proyecto combinó recursos proporcionados por la cátedra, incluyendo un prototipo básico de robot y materiales electrónicos, con la implementación de soluciones innovadoras realizadas por los autores, Maximiliano Exequiel Badano y Juan Fernando Asuar. Los docentes Néstor Balich y Franco Balich desempeñaron un rol de capacitación y mentoría con los cual los alumnos aplicaron conceptos de la catedra de robótica, investigaciones ya realizadas en el laboratorio, en base a ellos se montó el robot, la electrónica, la investigación de dos sistemas de visión artificial, conexión entre el mobile y el robot y la implementación de un sistema de visión artificial y los algoritmos de control para teleoperación y funcionamiento autónomo.

Este trabajo refleja una integración efectiva de recursos limitados con herramientas modernas de inteligencia artificial y robótica, logrando

un sistema funcional capaz de realizar tareas autónomas como el seguimiento de una pelota y su desplazamiento hacia un arco. Este enfoque representa una contribución significativa al aprendizaje práctico, alineándose con las exigencias académicas de la cátedra y las expectativas de futuras competencias, permitiendo la repetibilidad de la experiencia en modelo que permite escalarse a tareas y acciones más complejas que hacer al juego de fútbol autónomo.

1. INTRODUCCIÓN

La integración de inteligencia artificial (IA) en la robótica ha permitido avances significativos en la creación de sistemas autónomos capaces de interpretar su entorno y tomar decisiones en tiempo real [3]. Este trabajo se centra en el desarrollo de un robot móvil autónomo, diseñado para participar en una simulación de fútbol robótico, en la cual se requiere identificar y seguir una pelota, localizar un arco y ejecutar movimientos para empujar la pelota hacia el objetivo [4].

El objetivo principal de este proyecto fue demostrar cómo la implementación de algoritmos de visión artificial y control autónomo permite a los robots realizar tareas complejas con mínima intervención humana. Para ello, se integraron herramientas de programación avanzadas, como TensorFlow Lite para el procesamiento de imágenes, y se desarrolló una aplicación Android capaz de controlar el robot mediante comunicación Bluetooth [5].

Este proyecto representa un caso práctico del uso de IA para resolver problemas específicos en un entorno dinámico, destacando la relevancia de estas tecnologías en la mejora de la autonomía y la efectividad de los sistemas robóticos. Además, los resultados obtenidos contribuyen al conocimiento aplicado de la robótica y sientan las bases para su participación en futuras competencias de robótica móvil.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El desarrollo de este proyecto se llevó a cabo en tres etapas principales: investigación, construcción del robot y desarrollo del sistema de visión artificial y control.

Investigación Inicial: El primer paso consistió en analizar los requisitos para la creación de un sistema robótico autónomo capaz de participar en una simulación de fútbol robótico. Se revisaron diversas tecnologías disponibles, incluyendo microcontroladores, controladores de motores y bibliotecas de visión artificial como YOLOv3 [6], MobileNets [7] y EfficientNet [8]. Además, se exploraron diferentes enfoques para el procesamiento de imágenes, como el uso de modelos preentrenados y herramientas para la conversión de modelos de detección a formatos optimizados para dispositivos móviles.

Construcción del Robot: El robot se ensambló utilizando un microcontrolador ESP32S como unidad de control principal, un módulo controlador de motores L298N, dos motores de corriente continua, ruedas y un armazón metálico. Se realizaron ensayos para mejorar la localización y el mapeo del robot utilizando deep learning [9].

Los componentes electrónicos fueron configurados para proporcionar movilidad y capacidad de respuesta a los comandos enviados desde un dispositivo Android mediante comunicación Bluetooth.

Adicionalmente, se realizaron múltiples ensayos y calibraciones de los valores PWM de los motores. Este proceso fue crucial para garantizar un movimiento fluido y estable, ajustando la velocidad y dirección de las ruedas de forma precisa para responder a las instrucciones de control enviadas por la aplicación Android.

Desarrollo del Sistema de Visión Artificial y Control: El sistema de visión artificial fue diseñado utilizando TensorFlow Lite, con un enfoque en la detección de objetos en tiempo real [1], optimizando su ejecución en dispositivos móviles. La aplicación Android fue desarrollada en Android Studio utilizando Kotlin [2], integrando el modelo de detección para controlar al robot.

Se creó un dataset personalizado compuesto por una preclasificación de 700 imágenes de la pelota y el arco, capturadas desde diferentes ángulos y en diversas condiciones de iluminación. Este dataset fue utilizado para entrenar un modelo basado en YOLO, que se convirtió posteriormente a TensorFlow Lite mediante Google Colab para optimizar su ejecución en dispositivos móviles.

La aplicación Android, desarrollada en Android Studio utilizando Kotlin, integró este modelo de detección. Procesaba las imágenes capturadas por la cámara del dispositivo, identificaba la posición de los objetos clave (pelota y arco) y enviaba comandos al ESP32S a través de Bluetooth. El sistema de control del robot combinó esta información visual con algoritmos de toma de decisiones que le permitieron buscar, perseguir y empujar la pelota hacia el arco de manera autónoma. Este desarrollo implicó múltiples iteraciones para ajustar tanto el reconocimiento de objetos como la lógica de control, asegurando una integración efectiva entre la visión artificial y el hardware del robot Figura 1.

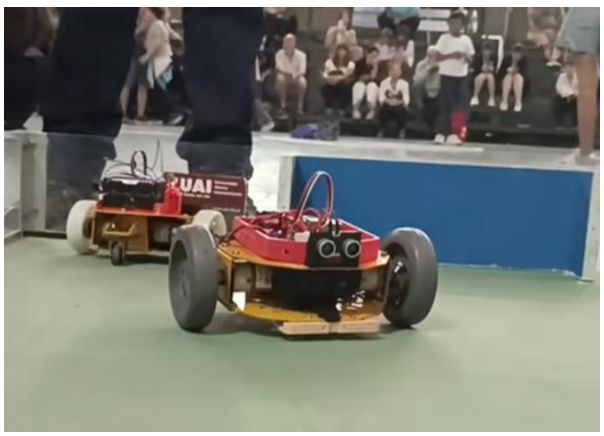


Figura 1. Robot con visión artificial.

3. RESULTADOS OBTENIDOS/ESPERADOS

Durante el desarrollo del proyecto, se lograron importantes avances en el diseño y funcionamiento del sistema robótico, aunque también se identificaron limitaciones inherentes a los recursos disponibles:

1. **Detección de objetos:** El modelo de visión artificial, entrenado únicamente para reconocer pelotas, mostró una alta eficiencia en la detección utilizando la cámara de un dispositivo Android. Las pruebas en un entorno controlado indicaron un desempeño robusto, aunque se reconoció que el modelo podría mejorarse aún más adicionando imágenes al dataset, especialmente con condiciones más variadas.
2. **Movimiento del robot:** El robot mostró capacidad de moverse hacia la pelota y orientarse hacia ella, aunque el desplazamiento no fue completamente fluido. Esto se debió a diferencias entre los motores utilizados, que generaron una ligera desviación hacia un lado a pesar de los esfuerzos de calibración.
3. **Optimización de instrucciones:** Se implementaron ajustes en el código de la aplicación Android para mejorar la coordinación entre las detecciones y los movimientos del robot, como ser la adición de tiempos de espera para evitar que el robot perdiera la pelota tras una detección fallida, permitiéndole reconocerla en intentos sucesivos.
4. **Pruebas en entorno controlado:** Aunque no se contó con un modelo entrenado para detectar el arco o una cancha delimitada, el robot cumplió con el objetivo de encontrar y seguir una pelota en condiciones controladas, mostrando un desempeño consistente y autónomo.

Para alcanzar una funcionalidad más avanzada y alineada con las expectativas de

competiciones futuras, se plantean las siguientes mejoras:

1. **Ampliación del dataset:** Entrenar el modelo con imágenes adicionales que incluyan el arco, la cancha y escenarios variados permitirá un reconocimiento más completo y robusto.
2. **Mejoras en hardware:** Reemplazar los motores actuales por un par de características idénticas reduciría las desviaciones en el movimiento, permitiendo un desplazamiento más fluido y estable.
3. **Optimización del sistema de control:** Continuar refinando la lógica de toma de decisiones en la aplicación Android y mejorar la sincronización entre detecciones e instrucciones enviadas al robot.
4. **Pruebas en entornos dinámicos:** Realizar pruebas en escenarios más complejos, incluyendo objetos en movimiento y variaciones de iluminación, permitirá evaluar la adaptabilidad y tiempo de respuesta del sistema.

Los resultados demostraron que el modelo de visión artificial tuvo una alta eficiencia en la detección de pelotas [6], aunque el desplazamiento del robot presentó limitaciones debido a diferencias en los motores. Se realizaron ajustes para mejorar la coordinación entre detecciones y movimientos, optimizando el sistema de control para smartphone [10] logrando reducir el costo del robot y permitiendo su repetibilidad para el resto de la comunidad de estudiantes e investigador.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo se encuentra conformado por dos estudiantes y dos docentes de la carrera de Ingeniería en Sistemas Informáticos de la Universidad Abierta Interamericana.

A lo largo de este proyecto, se adquirieron conocimientos en TensorFlow Lite, Kotlin y Android Studio, así como aplicaciones prácticas de diseño, robótica, Arduino y electrónica. Esta experiencia permitió fortalecer competencias en la integración de sistemas embebidos y en tecnologías de la información.

5. BIBLIOGRAFÍA

[1] TensorFlow documentation. TensorFlow. Recuperado el 20 de febrero 2025, de https://www.tensorflow.

[2] Android Developers documentation. Google. Recuperado el 20 de febrero 2025, de https://developer.android.com

[3] Cao, Y., Chen, Z., & Li, J. (2021). Integration of artificial intelligence in mobile robotics: Applications and challenges. Robotics and Autonomous Systems, 142, 103789. Recuperado de https://www.sciencedirect.com/science/article/pii/S0921889021000450

[4] Lee, S., & Kim, J. (2022). Vision-based control system for autonomous robots using deep learning. International Journal of Robotics Research, 41(2), 200–216. Recuperado de https://journals.sagepub.com/doi/10.1177/02783649211073320

[5] D. Pal, N. Kaur, R. Motwani, A. D. Mane and P. Pal, "Voice-Controlled Robot using Arduino and Bluetooth," *2023 3rd International Conference on Smart Data Intelligence (ICSMDI)*, Trichy, India, 2023, pp. 546-549, doi: 10.1109/ICSMDI57622.2023.00103. keywords: {Schedules;Bluetooth;Protocols;Microcontrollers;Process control;Robot sensing

systems;Software;Arduino;App;Bluetooth;Robot;Commands},

[6] Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. arXiv preprint arXiv:1804.02767. Recuperado de https://arxiv.org/abs/1804.02767

[7] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861. Recuperado de https://arxiv.org/abs/1704.04861

[8] Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. arXiv preprint arXiv:1905.11946. Recuperado de https://arxiv.org/abs/1905.11946

[9] Bazazian, D., Bartolozzi, C., & Censi, A. (2019). Robot localization and mapping using deep learning. IEEE Robotics and Automation Letters, 4(3), 2722–2729. Recuperado de https://ieeexplore.ieee.org/document/8712493

[10] "Smartphone Sensor-based Development and Implementation of a Remotely Controlled Robot Arm," JRC, vol. 5, no. 4, 2020. DOI: 10.18196/jrc.v5i4.21987.

Monitoreo continuo de Temperatura Hospitalaria Usando Sensores IoT

Departamento de Básicas y Aplicadas, Universidad Nacional de Chilecito 9 de Julio 22, Chilecito, La Rioja, Argentina [1]

Lic. Ariel Quiroga Marin[1], Ing. Enrique N. Martinez[1], Ing. Elvio Sigampa[1], Ing. Anzalaz Fernando[1], Ing. Gabriel Martinez[1], Dr. Cesar Salcedo[1], Hugo Fajardo[1,] Matías Perez[1] {cquiroga, emartinez, esigampa, fanzalaz, gmartinez, csalcedo, hfajardo, mperez}@undec.edu.ar, Andrés Marquez[1], Gastón Somerville[1] , Leonardo Zalazar[1] {aemb132, gastongg99, leonardozalazar89}@gmail.com

Hospital Eleazar Herrera Motta, Cavero 500, Chilecito, La Rioja[2]

Dra. Claudia Salguero[2], Dr. Omar Suarez[2], Leonel Alvarez[2], Rubén Quiroga[2]

{clausalguero2017}@yahaoo[2],{salguero, omarbreppe640, rubenhquiroga0, leonelundec}@gmail.com[2]

Resumen

El control preciso de la temperatura en entornos hospitalarios es crucial para mantener condiciones óptimas que aseguren la seguridad y el bienestar tanto de los pacientes, insumos médico y alimentación. Las importancias de la temperatura controlada en diversas áreas hospitalarias como la cocina y áreas de almacenamiento de alimentos, quirófanos, guardia central, nefrología, salas de pacientes adultos, pediátricos, neonatología, farmacia hospitalaria, mediante el monitoreo de temperatura y humedad basado en sensores IOT, puede jugar un papel crucial en la prevención en pacientes, contaminación de alimentos y medicamentos. Los sensores se encuentran en lugares estratégicos del hospital y transmiten los datos en tiempo real con el protocolo MQTT. Esta solución integrada de monitoreo de temperatura y humedad mediante sensores, puertas de enlace LoRaWAN[1] y plataforma de aplicaciones IOT (arduino cloud)[2] representan una herramienta poderosa para garantizar la calidad y seguridad de productos críticos, al tiempo que mejora la eficiencia operativa y el cumplimiento normativo en instituciones Sanitarias. Los dispositivos Internet of Things (IoT), no solo facilitan la gestión y la atención más eficiente, sino que también contribuyen significativamente a mejorar la seguridad y el bienestar de los pacientes, con el objetivo de alcanzar el "Riesgo Cero" en varios aspectos críticos de la atención sanitaria.

PALABRAS CLAVE: IOT, LoraWan, MQTT, Sensores

Líneas de I+D+I

- Procesamiento de Señales y Sistemas de Tiempo Real
- Arquitectura, Redes y Sistemas Operativos
- Base de Datos
- Cloud Computing
- Minería de Datos

Contexto

Los dispositivos de Internet de las Cosas (IoT), como las pulseras inteligentes y otros wearables, permiten el monitoreo continuo de los pacientes en tiempo real. Esto contribuye a reducir los tiempos de espera, optimizar la administración de medicamentos, gestionar eficientemente la farmacia hospitalaria y mejorar la calidad de la atención, con el objetivo de alcanzar un nivel de riesgo mínimo en el entorno clínico. Un ejemplo claro de esto es el uso de dispositivos IOT[3] como pulseras inteligentes que monitorizan en tiempo real las constantes vitales de los pacientes. Estos dispositivos, combinados con la conectividad 5G, permiten transmitir datos de forma instantánea a los profesionales de salud, facilitando

una respuesta rápida ante cualquier cambio crítico en el estado del paciente. Este tipo de tecnologías también está transformando la manera en que se brinda atención médica, haciendo los procesos más eficientes y seguros. El monitoreo de temperatura y humedad basado en IOT en El Hospital Eleazar Herrera Motta de Chilecito Fig1., no solo mejora la gestión operativa y la eficiencia logística, sino que también desempeña un papel crucial en la protección de la calidad de cuidado de los pacientes y seguridad de los alimentos y medicamentos a lo largo de toda la cadena de suministro. Esto es fundamental para asegurar que los productos lleguen a los consumidores finales en condiciones óptimas y seguras para su uso.



Fig1. Esquema representativo para monitorear y controlar sensores de temperatura en Hospital de Chilecito "Eleazar Herrera Motta" 29°10'08"S 67°29'15"W Chilecito (La Rioja)



Imágenes de UTI, Cocina y Neonatología Hospital Eleazar Herrera Motta Chilecito.

Introducción

El monitoreo de temperatura y humedad puede jugar un papel crucial en Salud hospitalaria y en la prevención en pacientes, contaminación de alimentos y medicamentos, también podemos monitorear con sensores IAQ integral que detectan concentración de CO2, temperatura, humedad, luz, nivel de HCHO/O3, TVOC, presión barométrica, PM2.5, PM10, Gas y movimiento. En nuestro trabajo no concentraremos en los siguientes aspectos:

1. Monitoreo en tiempo real: La tecnología LoRaWAN permite el monitoreo continuo y en tiempo real de la temperatura y la humedad en diferentes puntos del hospital Eleazar Herrera Motta (ejemplo: UTI, nefrología, alimentos y medicamentos). Esto es esencial para detectar cualquier desviación de las condiciones ideales de temperatura que podrían comprometer la calidad y seguridad de los pacientes , productos de farmacia y alimentos.

2. Alertas y notificaciones instantáneas: Los sensores pueden enviar alertas automáticas a los operadores o responsables cuando se detectan condiciones fuera de los rangos establecidos mediante MQTT (Message Queuing Telemetry Transport)[4]. Esto permite tomar acciones correctivas de manera rápida antes de que se produzcan daños o pérdidas significativas.

3. Registro de datos históricos: Además de la supervisión en tiempo real, se puede obtener el registro y almacenamiento de datos históricos de temperatura y humedad. Estos registros son cruciales para el análisis posterior, la trazabilidad y el cumplimiento de regulaciones sanitarias y de calidad.

4. Reducción de riesgos y costos: Al prevenir la contaminación de alimentos y medicamentos debido a condiciones ambientales inapropiadas, se reducen significativamente los riesgos de pérdidas económicas.

5. Cumplimiento normativo: En muchos sectores, especialmente en el ámbito de la salud y la alimentación, existen normativas estrictas sobre el

control de la temperatura y la humedad. Utilizar tecnología IOT para el monitoreo puede ayudar a cumplir con estas regulaciones de manera efectiva y eficiente.
El monitoreo de temperatura y humedad no solo mejora la gestión operativa y la eficiencia logística, sino que también desempeña un papel crucial en la protección de la calidad a los pacientes y seguridad de los alimentos y medicamentos a lo largo de toda la cadena de suministro. En nuestra investigación, realizamos simulaciones previamente de varios escenarios con el simulador Cupcarbon[5] para el sistema de detección y registro de temperatura, a través de sensores. Los sensores se encuentran en lugares estratégicos del hospital y transmiten los datos en tiempo real. La simulación se logra programando esos sensores en el lenguaje sen-script para una funcionalidad. El algoritmo comienza con la detección de temperaturas, que se genera con la ayuda de un generador de eventos naturales.
Para simular el comportamiento de los sensores utilizamos CupCarbon Digital Twin/IoT (herramienta de programación que se puede utilizar para diseñar y programar sistemas distribuidos, incluidas redes de sensores inalámbricos (WSN) e Internet de las cosas (IoT)), que nos permitirá simular comunicaciones mediante zigbee (IEEE 802.15.4), WiFi y LoRa. Se apoya en OpenStreetMaps lo que permite evaluar el comportamiento de las redes, permitendo a los usuarios utilizar el protocolo MQTT[6] para la comunicación de Internet de las cosas (IoT) Fig 2.

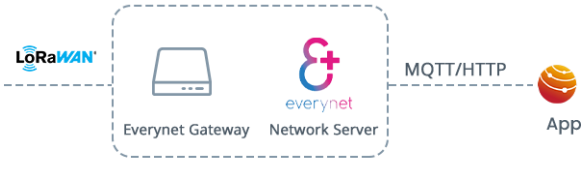




Fig 2. LORAWAN MQTT y Digital Twin/IoT Programming Cupcarbon

La conectividad inalámbrica y las tecnologías de sensores de IoT han fortalecido mutuamente la sostenibilidad y la innovación digital Fig. 3

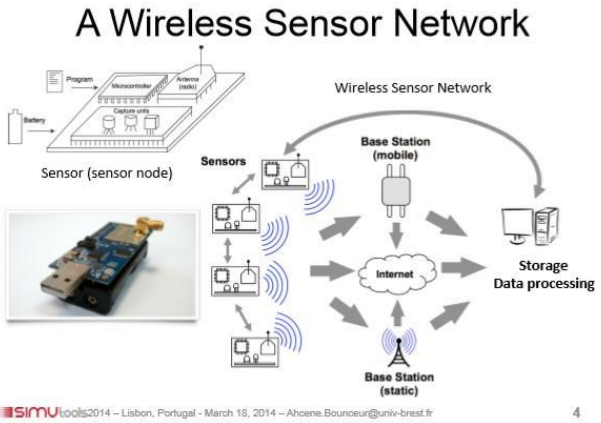


Fig 3. Red de sensores inalámbricos en IoT ayuda a aprovechar todo el potencial de la tecnología de IoT

La solución integrada de monitoreo de temperatura y humedad mediante sensores que lograremos con puertas de enlace LoRaWAN y plataforma de aplicación, representa una herramienta poderosa para garantizar la calidad y seguridad de productos críticos, al tiempo que mejora la eficiencia operativa y el cumplimiento normativo en diversas entidades de Salud[7] y la industria en general.

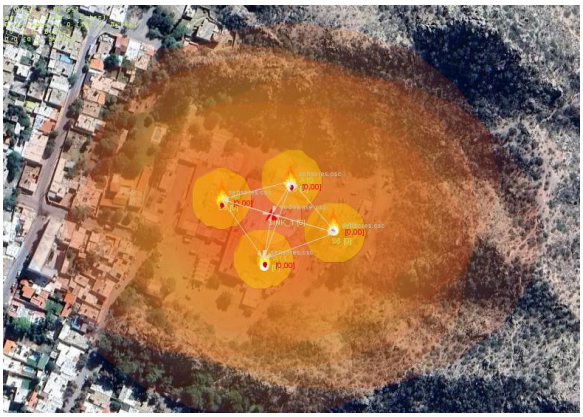


Fig 4. Simulación de sensores de temperatura Hospital de Chilecito con programación cupcarbon IOT Digital Twin/IoT

Los sensores IOT en las habitaciones de internados, cuidados intensivos, medicamentos, comida y otras áreas del hospital pueden controlar la temperatura, la humedad y la calidad del aire, creando ambientes más confortables y seguros para los pacientes, especialmente en unidades de cuidados intensivos o salas de aislamiento. Algunas aplicaciones claves de los dispositivos IoT en entornos hospitalarios son:

1. Monitorización continua de pacientes: Las pulseras inteligentes y otros wearables pueden medir constantemente parámetros vitales como el ritmo cardíaco, la presión arterial, la saturación de oxígeno y otros indicadores importantes. Esto permite a los profesionales de la salud supervisar el estado de los pacientes de manera remota y en tiempo real, lo que resulta especialmente útil en áreas como las unidades de cuidados intensivos y los postoperatorios.

2. Administración de medicamentos: Los dispositivos IOT pueden integrarse con sistemas de administración de medicamentos para asegurar que se administren las dosis correctas y en los horarios adecuados. Esto ayuda a prevenir errores de medicación y mejora la adherencia al tratamiento, proporcionando un nivel más alto de seguridad y eficacia.

3. Gestión de la farmacia hospitalaria: El monitoreo mediante IoT puede extenderse a la gestión de inventarios en farmacias hospitalarias, garantizando que los medicamentos se almacenen correctamente (como en heladeras de farmacia), se distribuyan eficientemente y se utilicen antes de su fecha de caducidad.

4. Seguridad en quirófanos, nefrología, cuidados intensivos, comida, farmacia, neonatología, vacunas e internados: La medición de la temperatura en estas áreas críticas es esencial para mantener condiciones óptimas que minimicen el riesgo de infecciones nosocomiales, alimentos, medicamentos, que aseguren la recuperación adecuada de los pacientes.

5. Mejora de la experiencia del paciente: Con el monitoreo continuo y la atención personalizada facilitada por los dispositivos IoT, los pacientes experimentan un cuidado más individualizado y menos intrusivo, lo cual puede reducir la ansiedad y mejorar su satisfacción general con la atención recibida.

Formación de Recursos Humanos

El equipo de trabajo está formado por docentes de las carreras Ingeniería en Sistemas, Licenciatura en Sistemas, Ing. Mecatrónica y Lic. En Enfermería de la UNdeC (acreditadas por CONEAU). También participan profesionales del Hospital Eleazar Herrera Motta de Chilecito, residentes y alumnos avanzados de grado en Mecatrónica, Enfermería y Sistemas, realizando su trabajo de tesina final en esta línea de I+D. Los integrantes son docentes de las asignaturas Sistemas Operativos I y II, Modelos y Simulación, Inteligencia Artificial, Seminario I, Taller Hardware e informática, Electrónica, Enfermería Hospitalaria. Estas asignaturas contemplan la aprobación mediante la participación en proyectos de investigación, por lo que pueden surgir nuevos trabajos en esta línea.

RESULTADOS ESPERADOS

Crear un entorno hospitalario inteligente[8] implica integrar tecnologías avanzadas para mejorar la eficiencia operativa, la atención al paciente y la productividad del personal médico. una solución integral y confiable de monitoreo de temperatura en hospitales es esencial para garantizar condiciones óptimas en áreas críticas como la cocina, los quirófanos, sala de urgencias, internados y la farmacia hospitalaria. Esto no solo promueve la seguridad y el bienestar de los pacientes y el personal, sino que también contribuye a cumplir con las normativas sanitarias y a mejorar la eficiencia operativa del establecimiento de Salud. La inclusión de la tecnología IOT smart city[9] en la sanidad no solo mejora la precisión y eficiencia de los procesos médicos, sino que también contribuye a una atención más segura y personalizada. Esta revolución digital está transformando la manera en que se gestionan los recursos y se proporciona la atención médica, promoviendo un entorno más conectado y orientado al paciente. En resumen, esta solución integrada de monitoreo de temperatura y humedad mediante sensores, puertas de enlace LoRaWAN y plataforma de aplicaciones arduino cloud, ESP32, MQTT, representan una herramienta poderosa para garantizar la calidad y seguridad de productos críticos, al tiempo que mejora la eficiencia operativa y el cumplimiento normativo en diversas industrias. Los dispositivos Internet of Things (IOT), no solo facilitan la gestión y la atención más eficiente, sino que también contribuyen significativamente a mejorar la seguridad y el bienestar de los pacientes, con el objetivo de alcanzar el "Riesgo Cero"[10] en varios aspectos críticos de la atención sanitaria. El monitoreo de la temperatura hospitalaria mediante sensores IoT mejora significativamente la seguridad, la eficiencia operativa y el cumplimiento de las normativas, al proporcionar datos en tiempo real y permitir una respuesta rápida ante cualquier desviación, esta tecnología contribuye a un ambiente hospitalario más seguro y controlado. Mantener una temperatura constante y controlada es esencial, si un sensor detecta que la temperatura se acerca a un rango crítico, el sistema genera una alerta que permite al personal tomar medidas inmediatas, para que esto funcione, los sensores IoT están ubicados estratégicamente en diversas zonas del hospital. Estos sensores monitorean la temperatura constantemente y transmiten la información en tiempo real a una plataforma central o al sistema de gestión hospitalaria.

Bibliografia, referencias

[1] Sabas. (2024, April 20). Haciendo IoT con LoRa: Capítulo 1.- ¿Qué es LoRa y LoRaWAN? Medium. https://medium.com/@Sabasacustico/haciendo-iot-con-lora-cap%C3%ADtulo-1-qu%C3%A9-es-lora-y-lorawan-8c08d44208e8

[2] Arduino Cloud | Build, Control, monitor your IoT Projects. (s. f.). https://cloud.arduino.cc/

[3] LoRaWAN backend Interfaces specification v1.0, LoRa Alliance de https://bit.ly/40RjZ8J

[4] Comunicación con aplicaciones (MQTT) https://lora-alliance.org/

[5] Design your Digital Twin IoT World https://cupcarbon.com/

[6] Comunicación con aplicaciones (MQTT) https://www.ibm.com/docs/es/masv-and-l/maximo-monitor/continuous-delivery?topic=reference-communicating-applications-mqtt

[7] Administración Nacional de Medicamentos, Alimentos y Tecnología Médica Disposición 3475/2005 https://www.argentina.gob.ar/normativa/nacional/disposici%C3%B3n-3475-2005-107238/texto

[8] Hospitales inteligentes (Organización Mundial de la Salud) https://www.paho.org/es/hospitales-inteligentes

[9] smart city (ciudad inteligente) https://es.wikipedia.org/wiki/Ciudad_inteligente

[10] Enfermería y seguridad de los pacientes Pag. 134 https://lildbi.fcm.unc.edu.ar/lildbi/tesis/9789275332467_spa.pdf

Sistemas de Tiempo Real, embebidos y simulaciones

Fernando Romero[1], Diego Encinas[1], Armando De Giusti[1,2], Lucas Maccallini, Alan Castelli[1], Horacio Villagarcía[1,3], Fernando G. Tinetti[1,3]

[1]Instituto de Investigación en Informática LIDI (III-LIDI)[3]
Facultad de Informática – Universidad Nacional de La Plata – Centro Asociado CIC

[2] CONICET – Consejo Nacional de Investigaciones Científicas y Técnicas.
[3] Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC)

{fromero, dencinas, degiusti, hvw, fernando}@lidi.info.unlp.edu.ar,
{lucas.maccallini, alanfabcast}@gmail.com

Resumen

Esta línea de investigación se dedica al desarrollo y análisis de Sistemas de Tiempo Real. Dentro de esta temática se trabajó sobre dos sublíneas: 1) Modelado y Simulación de Sistemas de Tiempo Real. 2) Sistemas de robótica y comunicaciones. Las simulaciones estudiadas comprenden circulación de personas en ambientes hospitalarios, transmisión de enfermedades y de virus, y circulación de pasajeros en trenes y aeropuertos y vehículos en autopistas. Los sistemas de robótica y comunicaciones vistos este año comprenden un robot equilibrado en 2 ruedas, un planificador de trayectorias de un robot, y un robot rastreador.

Contexto

Esta línea de Investigación forma parte del proyecto F032, que comprende Computación de Alto Desempeño y Distribuida: Arquitecturas, Algoritmos, Tecnologías y Aplicaciones en HPC, Fog-Edge-Cloud, Big Data, Robótica, y Tiempo Real del Instituto de Investigación en Informática LIDI acreditado por la UNLP.

Palabras Claves: Tiempo Real, Simulación, Comunicaciones, Robótica, Microcontroladores, Cloud Computing, Inteligencia Artificial, Reconocimiento de patrones.

1. Introducción

Para que un sistema sea considerado de Tiempo Real (STR) debe incluir restricciones de tiempo estrictas en los plazos en que, ante una determinada entrada, deben producir una salida. Cuentan como otra característica interactuar con el mundo físico, a través de entradas conectadas a sensores y detectores y salida a actuadores [1] [2] [3] [6] [8] [11] [13] [14] [17]. De acuerdo con el sistema a controlar serán las restricciones temporales. Se requiere sincronizar el tiempo de procesamiento con el tiempo físico. Para el procesamiento de estas entradas y salidas se utilizan placas basadas en microcontroladores como, Arduino, NodeMCU, CIAA [10] [18] [25] y Computadoras de Placa Simple como, Raspberry Pi, Raspberry Pi Zero

1

utilizando diferentes SOTR (Linux RT-Preempt, FreeRTOS, MQX, OSEK-OS, Zephyr, Raspbian, etc.) [4] [19] [27].

En el campo del modelado y simulación [8] [9] [12] [14] [15] [22] [26] [28] [29] , se trata de predecir el comportamiento y la eficiencia de distintos sistemas ante diferentes escenarios, siendo calibrados con datos reales.

2. Resultados y Objetivos

Se han desarrollado tareas sobre los temas antes expuestos tales como:

- Péndulo invertido: se desarrolló un sistema de balanceo autónomo de un vehículo de dos ruedas utilizando una plataforma ESP32. El sistema logra mantener el equilibrio de forma autónoma gracias a un algoritmo de control PID
- Planificador de trayectoria: se diseñó un algoritmo eficiente de planificación de trayectorias capaz de operar en tiempo real en el ESP32 y se implementó en un robot [12].
- Q-learning: Se construyó un Robot Rastreador que utiliza el algoritmo de aprendizaje Q-Learning para aprender a moverse mediante un brazo articulado con dos grados de libertad.
- Tráfico (Autopista): se desarrolló un modelo, basado en otro trabajo anterior de la Cátedra, sobre el tráfico en la Autopista 25 de Mayo de CABA[17].
- Trenes (La Plata): Se modeló el movimiento de trenes y pasajeros en el ferrocarril de La Plata [26][16].
- Transporte (Aviones): se modeló el tráfico aéreo nacional utilizando el software ARENA y basado en el trabajo previo de los compañeros del año pasado [5]
- Dengue: simulación diseñada para estudiar la propagación del virus del dengue y evaluar el impacto de la vacunación en una población [10][20][21]

3. Formación de Recursos Humanos

Se desarrollan trabajos de alumnos en la Convocatoria a Proyectos de Desarrollo e Innovación de la Facultad de Informática de la UNLP y de promoción de la Cátedra Sistemas de Tiempo Real de Ingeniería en Computación.

Se dirigen Prácticas Profesionales Supervisadas (PPS) con las que concluyen sus estudios los alumnos de Ingeniería en Computación y una tesina de la Licenciatura en Sistemas.

De postgrado, investigadores del grupo están desarrollando un trabajo final de especialización, una tesis de Maestría y una tesis de Doctorado.

4. Referencias

[1] Integration of Sensor Networks with Cloud ComputingS. Medina, F. Romero, and F. G. Tinetti, Short papers of the 8th Conference on Cloud Computing, Big Data & Emerging Topics (JCC-BD&ET 2020), ISBN: 978-950-34-1927-4, págs. 2-5, 2020.

[2] Buttazzo, G. C., "Hard RealTime Computing Systems", Third edition, Springer, 2011.

[3] Burns, A, A, Wellings. "Real-Time Systems and Programming Languages: Ada, Real-Time Java and C/Real-Time POSIX", Addison-Wesley Educational Publishers Inc., 2009.

[4] FreeRTOS - market leading RTOS (real time operating system) for embedded systems supporting 34 microcontroller architectures. http:// www.freertos.org/.

[5] https://www.rockwellautomation. com/esmx/products/software/arena-simulation.html MinisteriodeTransporte. Agencia Nacional de AviaciónCivil. Recuperado de https://www.argentina. gob.ar/anac (Octubre,2023)].

[6] Kopetz. H., "Real-Time Systems, Design Principles for Distributed Embedded Applications". Second Edition. Springer. 2011.
[7] Liu, J. W. S. Liu, "Real Time Systems", Integre Technical Publishing Co., Inc., 2000
[8] C. Macal, M. North, Tutorial on agent-based modeling and simulation part 2: how to model with agents, in: Proceedings of the Winter Simulation Conference, 2006.
[9] PHILLIP A. LAPLANTE, SEPPO J. OVASKA. REAL-TIME SYSTEMS DESIGN AND ANALYSIS Tools for the PractitionerFourth Edition. A JOHN WILEY & SONS, INC., PUBLICATION. IEEE PRESS. 2012.
[10] Chin,Pamela&Ehrbar,Dylan & Micieli,María&Fonseca,Dina&Ciota, Alexander& Kramer,Laura.(2018). "DifferentialEffectsofTemperatureandMosquitoGenetics DetermineTransmissibilityofArboviruses byAedesAegyptiinArgentina".TheAmerican JournalofTropicalMedicine andHygiene .99.10.4269/ajtmh.18-0097.
[11] R.S. Rosli, M.H. Habaebi, R. Islam On the analysis of received signal strength indicator from ESP8266, Bulletin of Electrical Engineering and Informatics · September 2019
[12] Análisis de una plataforma de simulación para Cloud Computing. Un caso de estudio. Tomás Rosales, Julián Spinelli, Marcos Di Nardo, Román Bond, Daniel Rosatto, Diego Encinas, Fernando Romero XXVI Congreso Argentino de Ciencias de la Computación (CACIC 2020) La Matanza
[13] F. G. Tinetti and O. C. Valderrama Riveros, "Unmanned Vehicles: Towards Heterogeneous Hardware Approaches," 2018 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 2018, pp. 919-924.
[14] Fernando G. Tinetti, Oscar C. Valderrama Riveros, Fernando L. Romero, "Unmanned Vehicles: Real Time Problems in Drone Receivers", Conf. on Computational Science & Computational Intelligence (CSCI'19), Las Vegas, Nevada, USA , 2019, pp. 1081-1085.
[15] D. Black, SystemC: From the Ground Up. Second Edition, Springer, 2010.
[16] Ministerio de Transporte Argentino, Cantidad de transacciones por fecha de la tarjeta SUBE(2020-2023) https:// datos. gob.ar/ dataset/transporte- sube — cantidad- transacciones- usos-por- fecha
[17] Buenos Aires Data, Flujo Vehicular por Unidades de Peaje AUSA, URL: https://data.buenosaires.gob.ar/dataset/fluj o-vehicular-por-unidades-de-peaje-ausa].
[18] NodeMcu http://www.nodemcu.com/ 2017
[19] https://www.ros.org/.
[20] Maccallini, L. ., Encinas, D. O., & Romero, F. . (2021). An Approach to the Modeling and Simulation of Intra-Hospital Diseases. Journal of Computer Science and Technology, 21(2), e14. https://doi.org/10.24215/16666038.21.e14
[21] Encinas, D. O., Matto, J., Stella, J., Maccallini, L., & Romero, F. (2024). Dengue Dynamics: Modelling Spread and Environmental Interactions. Journal of Computer Science and Technology, 24(2), e14. https://doi.org/10.24215/16666038.24.e14
[22] Uri Wilensky (29 de septiembre de 2022). NetLogo User Manual version 6.3.0. Disponible en Internet en: https://ccl.northwestern.edu/netlogo/docs/
[26] M. Paradiso, L. Maccallini, A. Vericat, F. Romero and D. Encinas, "An approach to parallelization of respiratory disease spread simulations in emergency rooms," in 2022 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 2022 pp. 1359-1365. doi: 10.1109/CSCI58124.2022.00244.

[27] Mauro Santos, Diego Encinas. Análisis de plataformas de Computación en la Nube para implementación de protocolo de comunicaciones con una aplicación móvil 3D. XXVIII Congreso Argentino de Ciencias de la Computación (CACIC 2022) La Rioja
[28] D. Encinas, S. Mendez, M. Naiouf, A. De Giusti, D. Rexachs del Rosario, and E. Luque, "An Agent-Based Model for Analyzing the HPC Input/Output System" International journal on advances in systems and measurements vol. 13, num. 3 & 4, págs. 192-202, 2020.
[29] D. Encinas, M. Naiouf, A. De Giusti, S. Mendez, D. Rexachs del Rosario, and E. Luque, "On the calibration, verification and validation of an agent-based model of the hpc input/output system" In Proceedings from The Eleventh International Conference on Advances in System Simulation (SIMUL 2019), 2019.

ORCID autores:
Fernando Romero: 0000-0002-1498-3752
Diego Encinas: 0000-0002-6948-9786
A. De Giusti: 0000-0002-6459-3592
Santiago Medina: 0000-0001-6852-7165
Horacio Villagarcía:
Fernando G. Tinetti:
César Estrebou: 0000-0001-5926-8827

Telemetría para el Monitoreo Ambiental en la cámara de fructificación de cultivo de hongos en INTA Mendoza

Matilde Inés Césari[1], Cecilia Césari[2] Valeria Espartaro[1], María Cecilia Castellanos[1], Eugenia Alfonso[1], Carolina Pennisi[1], Noelia Robles[1,] María Eugenia Stefanoni[1], Daniela Carbonari,[1]

[1]GIDECo Grupo Regional UTN de Investigación y Desarrollo de Ecosistemas de Conocimiento - ReAVi Laboratorio de investigación Realidad Aumentada y Virtual) – Universidad Tecnológica Nacional FRM / matilde.cesari@frm.utn.edu.ar

[2]INTA-EEM Estación Experimental Mendoza INTA / cesari.cecilia@inta.gob.ar

RESUMEN

El objetivo de este estudio es diseñar, desarrollar e implementar un sistema de control y monitoreo electrónico basado en tecnología Arduino y plataformas IoT para la cámara de fructificación del cultivo de hongos en INTA. Se recopilaron datos críticos de temperatura, humedad y CO2, los cuales se transmitirán a la plataforma ThingSpeak para su monitoreo en tiempo real. Los resultados muestran la precisión y eficiencia del sistema en el mantenimiento de condiciones ambientales óptimas, minimizando el error humano y estableciendo las bases para un sistema de control automatizado.

Palabras clave: Telemetría, monitoreo ambiental, IoT, cultivo de hongos.

INTRODUCCIÓN

La adopción de tecnologías avanzadas en la agricultura moderna, particularmente en el cultivo de hongos, juega un papel crucial en la mejora de la eficiencia y la productividad. Los sistemas de monitoreo y control ambiental automatizados son esenciales para mantener las condiciones óptimas de fructificación, ya que los hongos son extremadamente sensibles a las variaciones de temperatura, humedad, CO2 y otros factores ambientales La producción de Pleurotus ostreatus puede ser abordada desde una perspectiva sustentable (Miranda María, Gutiérrez-Segura, Cedillo, & Plata, 2023).

Estos sistemas permiten un manejo preciso y constante, lo que reduce la dependencia del monitoreo manual y minimiza los errores humanos. Además, el uso de sensores y tecnologías de IoT facilita la recopilación de datos en tiempo real, permitiendo ajustes inmediatos en las condiciones del cultivo y mejorando la calidad y el rendimiento de la producción de hongos (Kathiria *et al.*, 2023; Seetha Rama, 2021).

El objetivo específico de este trabajo es diseñar, desarrollar e implementar un sistema de control y monitoreo electrónico basado en tecnología Arduino y plataformas IoT para la cámara de fructificación del cultivo de hongos Pleurotus Ostreatus, garantizando el mantenimiento de condiciones ambientales óptimas. Este proyecto se ha desarrollado en colaboración entre la Estación Experimental Agropecuaria (EEA) Luján de Cuyo del Instituto Nacional de Tecnología Agropecuaria (INTA) y la Facultad Regional Mendoza de la UTN.

Se han realizado diversos estudios que destacan la importancia de la telemetría y el monitoreo ambiental en la agricultura, especialmente en el cultivo de hongos Pleurotus. Por ejemplo, el uso de sistemas de monitoreo ambiental basados en IoT en granjas de hongos ha demostrado ser efectivo para mantener condiciones óptimas de temperatura y humedad, mejorando la calidad y cantidad de la producción. Kassim, Mat y Yusoff (2019) describen cómo la implementación de sensores de IoT en granjas de hongos puede

monitorear y controlar de manera eficiente las condiciones ambientales, resultando en un manejo más preciso y una mayor productividad. Estudios como los de Mahmud *et al.* (2018) han mostrado que los sistemas basados en IoT pueden proporcionar monitoreo en tiempo real de variables críticas como temperatura y humedad, asegurando que las condiciones permanezcan dentro de los rangos óptimos para el crecimiento del hongo Pleurotus ostreatus.
Estos sistemas no solo reducen la dependencia del monitoreo manual, sino que también permiten ajustes automáticos que mejoran la eficiencia y reducen el riesgo de pérdida de cultivo debido a condiciones ambientales inadecuadas.

Adicionalmente, Kencanawati (2017) destacó la influencia de las condiciones ambientales en el cultivo ecológico de hongos Pleurotus en Indonesia, subrayando la necesidad de un monitoreo constante para asegurar un entorno de cultivo favorable. La investigación de Rahmat *et al.* (2020) demostró que la adición de ciertos aditivos al medio de cultivo, monitoreados mediante sistemas IoT, puede optimizar el crecimiento y la producción de hongos Pleurotus ostreatus.

En conjunto, estos estudios muestran que la integración de tecnologías IoT en el cultivo de hongos Pleurotus no solo optimiza las condiciones de cultivo y mejora la eficiencia, sino que también puede aumentar la producción y calidad de los hongos, proporcionando beneficios significativos tanto para los productores como para los consumidores.

METODOLOGÍA

La implementación del código se centró en la recolección de datos ambientales críticos utilizando los sensores seleccionados y la transmisión de estos datos a la plataforma ThingSpeak para su monitoreo remoto y en tiempo real.
Esta tecnología se enmarca dentro de las aplicaciones de IoT en la agricultura, la cual ha sido reconocida por su capacidad de mejorar la precisión y la eficiencia en el manejo de variables ambientales (Babakhouya *et al.*, 2023).

A continuación, se describe un ejemplo simplificado del proceso implementado:

1. **Inicialización:** Esta etapa involucra configurar el entorno de trabajo, inicializar las bibliotecas necesarias para los sensores y la conectividad WiFi en el controlador Wemos D1 mini. Se establece la conexión a la red WiFi y se configuran los pines digitales que leerán los datos de los sensores. El código comienza estableciendo la conexión WiFi utilizando las credenciales de la red (SSID y contraseña). Luego, inicializa la comunicación con los sensores DS18B20 (temperatura), DHT11 (humedad), y MQ7 (dióxido de carbono) configurando los pines correspondientes y las bibliotecas necesarias. Este paso es crucial para garantizar que todos los componentes estén listos para operar antes de recoger datos.
2. **Recolección de Datos:** Se leen los datos de los sensores a intervalos regulares. Por ejemplo, para el sensor de temperatura DS18S20, se solicita la temperatura actual; para el DHT11, se leen tanto la temperatura como la humedad; y para el MQ135, se mide la calidad del aire. En el loop principal, el sistema recopila datos de temperatura ambiente y humedad con el sensor DHT11, la temperatura de la bolsa de hongo con el sensor DS18B20, y los niveles de CO2 con el sensor MQ7. Estas lecturas se realizan en intervalos regulares definidos por el tiempo de espera al final del loop (delay(20000)).
3. **Procesamiento de Datos:** Los datos recogidos se procesan para adecuarlos a los formatos necesarios, aplicando, si es necesario, correcciones por calibración. Esto puede incluir la conversión de señales analógicas a digitales, el ajuste de las lecturas según factores de escala, y la preparación de los datos para su transmisión. Aunque el código adapta los valores leídos a formatos adecuados para su transmisión. Por ejemplo, convierte las lecturas de temperatura y humedad a cadenas de

caracteres para su visualización y prepara los datos para su envío a ThingSpeak.

4. **Transmisión a ThingSpeak:** Los datos procesados se envían a ThingSpeak utilizando la API de la plataforma. Esto generalmente se hace mediante una petición HTTP POST, incluyendo los valores de los sensores como parte del cuerpo de la petición. Cada sensor puede ser asignado a un campo diferente en un canal de ThingSpeak. Utiliza la API de ThingSpeak para enviar los datos recolectados. Cada valor de los sensores se asigna a un campo específico en un canal de ThingSpeak utilizando ThingSpeak.setField(). La conexión al servidor de ThingSpeak se realiza mediante cliente.connect(), y luego se envían los datos con ThingSpeak.writeFields(). Este proceso se verifica con el código de estado HTTP, indicando si la actualización fue exitosa. Utiliza la API de ThingSpeak para enviar los datos recolectados. Cada valor de los sensores se asigna a un campo específico en un canal de ThingSpeak utilizando ThingSpeak.setField(). La conexión al servidor de ThingSpeak se realiza mediante cliente.connect(), y luego se envían los datos con ThingSpeak.writeFields(). Este proceso se verifica con el código de estado HTTP, indicando si la actualización fue exitosa.
5. **Monitoreo Remoto**: Una vez que los datos están en ThingSpeak, se pueden visualizar en tiempo real a través de su interfaz web. Los usuarios pueden configurar alertas o acciones basadas en ciertos umbrales de datos, permitiendo una respuesta rápida a las condiciones cambiantes del entorno de cultivo. Al enviar los datos a ThingSpeak, habilita la visualización y análisis de los datos en tiempo real a través de la plataforma ThingSpeak. Los usuarios pueden acceder a estos datos desde cualquier lugar, permitiendo un seguimiento continuo de las condiciones ambientales de la cámara de fructificado.

RESULTADOS

Los resultados obtenidos durante las pruebas del sistema mostraron que las mediciones de temperatura, humedad y CO2 eran precisas y consistentes con los valores esperados. Las gráficas generadas en ThingSpeak proporcionaron una visualización clara y en tiempo real de las condiciones ambientales en la cámara de fructificación, destacando la efectividad del uso de tecnologías IoT en la agricultura de precisión (Vijayan *et al.*, 2024).

- Gráfica de Temperatura (Figura 2): La temperatura medida mostró fluctuaciones diarias, manteniéndose dentro del rango óptimo para la fructificación de 16-18°C. La gráfica adjunta muestra las variaciones de temperatura a lo largo de 14 día (una oleada).
- Gráfica de Humedad (Figura 3): La humedad relativa se mantuvo durante los primeros 7 días, dentro del rango óptimo para la formación de primordios de 95-98%. En los 7 días siguientes, se mantuvo dentro del rango óptimo para el crecimiento de los cuerpos (fructificación) de 85 a 95%. La gráfica adjunta proporciona una visión clara de cómo varía la humedad en el ambiente controlado.
- Gráfica de CO2 (Figura 4): Los niveles de CO2 se mantuvieron estables alrededor de 400-420 ppm en los primeros 10 días, luego fue incrementando hasta 1000ppm, asegurando una buena ventilación y calidad del aire en la cámara de fructificación. La gráfica adjunta ilustra los cambios en las concentraciones de CO2 durante un ciclo diario.

Las variaciones en la temperatura afectan directamente la tasa de crecimiento del micelio y la producción de frutos. Temperaturas fuera del rango óptimo pueden ralentizar el crecimiento, reducir el rendimiento y afectar la calidad de los hongos. Además, cada especie tiene sus propias preferencias y rangos óptimos, por lo que es crucial ajustar las condiciones ambientales según la especie específica de Pleurotus que se esté cultivando.

Figura 2. Mediciones de Temperatura en la Cámara de Fructificación

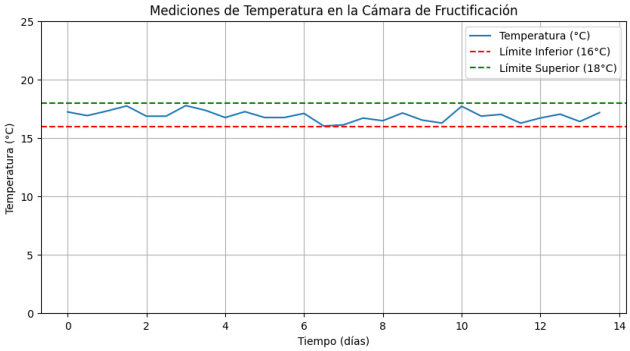


Figura 3 Mediciones de Humedad en la Cámara de Fructificación

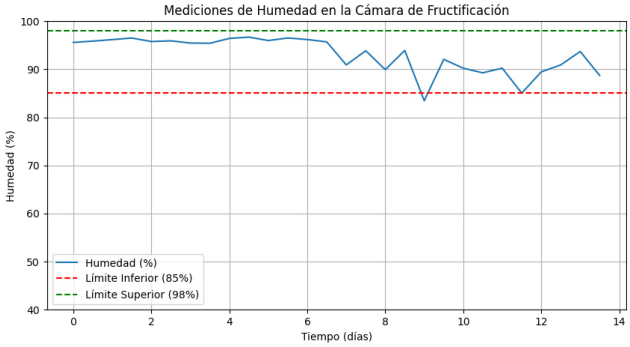
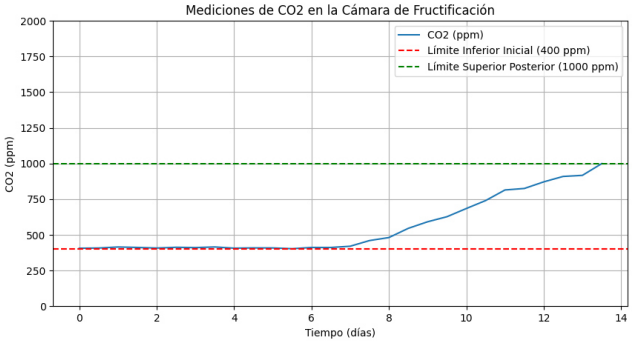


Figura 4. Mediciones De CO2 En La Cámara De Fructificación



DISCUSIÓN

En nuestra zona, la garantía de condiciones ambientales óptimas desempeña un papel importante en la optimización de la producción de gírgolas. A nivel general, la temperatura y humedad óptimas para una sala de fructificación de gírgolas son de alrededor de 18 a 20 °C y 85 a 95%, respectivamente (Díaz Muñoz *et al.*, 2019; Sánchez & Royse, 2001; Miranda *et al.* 2023). En el presente estudio, los resultados sugieren que se logró el establecimiento exitoso del sistema de monitoreo ambiental para el cultivo artificial de gírgolas tipo Pleurotus ostreatus ofreciendo una solución prometedora para aumentar el rendimiento de hongos.

Las ventajas del sistema incluyen la reducción del error humano y la capacidad de realizar ajustes inmediatos basados en datos en tiempo real.

Sin embargo, se identificaron limitaciones, como la necesidad de calibrar periódicamente los sensores para mantener su precisión y la dependencia de una conexión estable a Internet para la transmisión de datos. Este enfoque de monitoreo también ha sido respaldado por investigaciones que destacan la necesidad de una vigilancia constante para mantener la eficiencia en el cultivo de hongos (Khusnul, 2020).

Es importante destacar que estos resultados son parte de una etapa preliminar enfocada en la recopilación de datos antes de la implementación de un sistema remoto en un servidor web con funciones de control avanzadas.

Las mejoras futuras incluyen la integración de un sistema de respaldo para la transmisión de datos en caso de fallos de conexión y la implementación de algoritmos de control más avanzados, como la lógica difusa, para una gestión más precisa de las variables ambientales.

CONCLUSIONES

El sistema de telemetría implementado demuestra ser una herramienta efectiva para el monitoreo y control ambiental en la cámara de fructificación del cultivo de hongos. La recopilación y transmisión de datos en tiempo real permite una gestión precisa y eficiente de las condiciones ambientales, reduciendo el error humano y mejorando la calidad y rendimiento del cultivo. La colaboración entre INTA y UTN ha sido clave en el desarrollo de este sistema, que representa un avance significativo en la automatización del cultivo de hongos.

REFERENCIAS

Babakhouya, A., Naji, A., Hnini, A., & Daaif, A. (2023, May). Agricultural IoT technology: An overview of usages, technologies, and challenges. In *2023 3rd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)* (pp. 1-8). IEEE. https://ieeexplore.ieee.org/abstract/document/10153038

Díaz Muñoz, K., Casanova Guajardo, M., León Torres, C. A., Gil Ramírez, L. A., Bardales Vásquez, C. B., & Cabos Sánchez, J. (2019). Producción de *Pleurotus ostreatus* (Pleurotaceae) ICFC 153/99 cultivado sobre diferentes residuos lignocelulósicos. *Arnaldoa, 26*(3), 1177-1184. https://www.researchgate.net/publication/348406173_Produccion_de_Pleurotus_ostreatus_Pleurotaceae_ICFC_15399_cultivado_sobre_diferentes_residuos_lignocelulosicos

Islam, T., Zakaria, Z., Hamidin, N., & Ishak, M. A. B. M. (2016). Optimization of humidifying procedure in controlled environment for indoor cultivation of *Pleurotus pulmonarius*. *African Journal of Biotechnology, 15*(45), 2578-2586. https://www.ajol.info/index.php/ajb/article/view/148022

Kassim, M. R., Mat, I., & Yusoff, I. M. (2019). Applications of Internet of Things in mushroom farm management. *13th International Conference on Sensing Technology (ICST)* (pp. 1-6). Sydney, NSW, Australia. https://www.researchgate.net/publication/340238033_Applications_of_Internet_of_Things_in_Mushroom_Farm_Management

Kathiria, A., Barot, P., Paliwal, M., & Shastri, A. (2023, June). IoT-assisted mushroom cultivation in agile environment. In *International Conference on Data & Information Sciences* (pp. 299-308). Springer Nature Singapore. https://www.researchgate.net/publication/377074641_IoT-Assisted_Mushroom_Cultivation_in_Agile_Environment

Kencanawati, I. (2017). The influence of environmental conditions on ecological farming and production of oyster mushroom (*Pleurotus ostreatus*) in Kerinci, Indonesia. *The International Journal of Science & Technoledge, 5*(1), 70-73. https://www.internationaljournalcorner.com/index.php/theijst/article/view/123420/84582

Khusnul, K. (2020). Training and feasibility study of oyster mushroom cultivation at Ganda Mandiri Farmer Group in Sukamanah Village, Ciamis Regency. *Jurnal Pengabdian kepada Masyarakat (Indonesian Journal of Community Engagement), 8*(2), 102-106. https://journal.ugm.ac.id/jpkm/article/view/50748

Mahmud, M. S., Buyamin, S., Mokji, M. M., & Abidin, M. Z. (2018). Internet of Things based smart environmental monitoring for mushroom cultivation. *Indonesian Journal of Electrical Engineering and Computer Science, 10*(3), 847-852. https://doi.org/10.11591/ijeecs.v10.i3.pp847-852

Miranda María, G., Gutiérrez-Segura, E., Cedillo, J., & Plata, M. (2023). Análisis de producción de *Pleurotus ostreatus* desde un enfoque sustentable. In *La esfera de la vida: De la epistemología ambiental a la inter y transdisciplinariedad*. Editorial Torres Asociados. https://www.researchgate.net/publication/378309655_ANALISIS_DE_PRODUCCION_DE_PLEUROTUS_OSTREATUS_DESDE_UN_ENFOQUE_SUSTENTABLE

Rahmat, R., Rahim, I., Putera, M., & Asrul, L. (2020). Growth and production of white oyster mushroom (*Pleurotus ostreatus*) by adding coconut water to agricultural waste as a carbon source media. *IOP Conference Series: Earth and Environmental Science, 575*, 012090. https://doi.org/10.1088/1755-1315/575/1/012090

Sánchez, J., & Royse, D. (2001). *La biología y el cultivo de Pleurotus spp*. Uteha (Noriega Editores)-Ecosur. https://www.researchgate.net/publication/256526787_Book_La_biologia_y_el_cultivo_de_Pleurotus_spp

Seetha Rama, V. (2021). Precision agriculture using IoT. *International Journal for Research in Applied Science and Engineering Technology, 9*, 122-127. https://doi.org/10.22214/IJRASET.2021.36255

Vijayan, B., Haleem, A., Irbaz, A., Rehman, F., & Pasha, F. (2024).. Wireless networking: Smart agriculture. *Journal of Computer Science and Policy Integration, 3*(1), 1-8. https://doi.org/10.46610/jcspic.2024.v03i01.005

· PDP ·

Procesamiento Distribuido y Paralelo

Abordajes en modelado y simulación en sistemas de HPC y salud.

Diego Encinas[1,2], Jimena Jara[1], Román Bond[1], Daniel Rosatto[1], Diego Alvarez[1], Gustavo García Krahn[1], Ignacio Di Loreto[1], Verónica Scholz[1] Adriana Gaudiani[3], Martin Morales[1,4]

[1]Proyecto de Investigación SimHPC - Programa TICAPPS - Instituto de Ingeniería y Agronomía - UNAJ

[2]Instituto de Investigación en Informática LIDI (III-LIDI) - Facultad de Informática - UNLP – Centro Asociado CIC

[3]Área de Computación - Instituto de Ciencias - UNGS

[4]Centro CodApli - Facultad Regional La Plata - UTN

dencinas@unaj.edu.ar, elchejime@gmail.com, rbond@unaj.edu.ar, drosatto@unaj.edu.ar, diegoalvarez3219@gmail.com, krahn_gustavo@hotmail.com, diloretoignacio@gmail.com, vero.scholz@gmail.com, agaudiani@ungs.edu.ar, martin.morales@unaj.edu.ar

Resumen

El objetivo de esta línea de investigación es el estudio de diferentes técnicas de modelado y simulación para entornos de Cómputo en Altas Prestaciones (HPC, High Performance Computing). El enfoque del estudio es la obtención de herramientas que permitan predecir la eficiencia del sistema ante posibles escenarios y reconfigurar el sistema físico. Además, se analizan los diferentes componentes del sistema que pueden influir en las prestaciones significativamente y pueden llegar a modelarse y/o reconfigurarse.

Por otra parte, se ha llevado a cabo el desarrollo de un simulador para entornos en el área de salud, en el sector de emergencias hospitalarias y de propagación de enfermedades.

Palabras clave: *Arquitecturas Multiprocesador. Simulación. Sistema E/S paralela. Modelado y Simulación basado en agentes (Agent-Based Modeling and Simulation, ABMS). Cloud Computing. CloudSim. Simulación y Salud.*

Contexto

Se presenta una línea de Investigación que es parte del Proyecto de Investigación "Simulación, Computación de Altas Prestaciones (HPC) y optimización de aplicaciones sociales – SimHPC" de la Universidad Nacional Arturo Jauretche (UNAJ), acreditado por resolución interna 183/21. Además, el proyecto aporta al Programa "Tecnologías de la información y la comunicación (TIC) en aplicaciones de interés social" – TICAPPS de la UNAJ.

En el tema existe un convenio de colaboración en actividades de Investigación y Postgrado con el Instituto de Investigación en Informática – LIDI de la Universidad Nacional de La Plata y el Área de Computación del Instituto de Ciencias de la Universidad Nacional de General Sarmiento.

Introducción

El crecimiento sostenido en la demanda del poder de cómputo remarca la necesidad de sistemas con enfoques de paralelización masiva y cómputo de alta performance (HPC, High Performance Computing) [1]. Los clusters se han convertido en uno de los enfoques principales para lograr paralelismo a bajo costo. Una noción extendida lo constituye la utilización de grid computing y más recientemente cloud computing. Independientemente de la solución, estos sistemas constan de un gran número de componentes incluyendo nodos de procesamiento, bancos de memoria, discos, entre otros.

En cuanto a las herramientas de simulación para Cloud Computing, CloudSim es un framework desarrollado en Java que provee las APIs necesarias para que el usuario genere una simulación de un Data

Center funcionando como servidor de nube capaz de simular la ejecución de CloudLets.

Por otra parte, los servicios de urgencias hospitalarias son considerados como una de las unidades del sistema sanitario de mayor complejidad y fluidez, lo que unido a la variabilidad de su actividad da lugar a que su gestión operativa sea una tarea muy complicada [2]. Es por ello que resultaría de mucha utilidad para sus responsables disponer de un sistema de ayuda a la toma de decisiones (Decision Support System-DSS) tan flexible como un simulador, que permitiría tomar medidas disponiendo de información suficiente sobre las alternativas posibles.

Sistemas de E/S Paralela

Las exigencias en los sistemas de E/S paralelos se han incrementado debido al aumento en número, velocidad y potencia de las unidades de procesamiento en los clusters. También las aplicaciones científicas que utilizan cómputo de altas prestaciones acrecientan estos requerimientos.

En muchos casos, el cuello de botella de los sistemas paralelos es la E/S a causa de las exigencias que debe afrontar [3]. La E/S Paralela es esencial para emparejar el avance de las arquitecturas de los procesadores y el rápido crecimiento de la capacidad computacional. Aunque la arquitectura jerárquica de memoria multinivel puede evitar grandes pérdidas de prestaciones debido a los retardos de acceso a disco, la capacidad de memoria es limitada. Además, como la capacidad computacional aumentará, la disponibilidad de memoria por core decrecerá, especialmente si la escala de los sistemas de HPC se proyecta a millones de cores o más. Varias simulaciones científicas y de ingeniería de áreas críticas de investigación, tales como la nanotecnología, astrofísica, clima y energía física están convirtiéndose en aplicaciones intensivas de datos. Para poder disminuir la brecha entre CPUs-E/S se deben identificar los factores que influyen en las prestaciones y proponer nuevas soluciones [4] [5].

En el área de tolerancia a fallas en sistemas de cómputo de alta prestaciones se puede notar la importancia de la unidad de E/S en las arquitecturas paralelas como un punto a mejorar para lograr cubrir las exigencias de las aplicaciones que utilizan HPC.

Se puede disminuir la complejidad y la probabilidad de errores en la generación de sistemas híbridos desarrollando una simulación específica de éstos utilizando diferentes frameworks [6] [7] [8].

Las aplicaciones científicas con un uso intensivo de datos utilizan software de E/S paralelo para acceder a archivos. Contar con una herramienta que permita predecir el comportamiento de este tipo de aplicaciones en HPC es de gran utilidad para los desarrolladores de aplicaciones paralelas como para administradores de centros de cómputo. Por otro lado, ABMS ha sido utilizado para modelar problemas y sistemas complejos en diversas áreas de la ciencia.

Evaluar las prestaciones del subsistema de E/S con diferentes configuraciones y la misma aplicación, permite adaptar la configuración de E/S teniendo en cuenta el patrón de acceso de la aplicación. Pero también puede ser una gran ventaja analizar las necesidades de las aplicaciones antes de configurar el sistema físico. Una manera de predecir el comportamiento de las aplicaciones en el sistema de cómputo, ante distintas configuraciones, es utilizando técnicas de modelado y simulación.

Se está desarrollando modelos e implementando una simulación de la arquitectura de E/S paralela, por medio de técnicas de simulación basadas en agentes o Sistemas Multi-Agente, (MAS-MultiAgentSystems), para evaluar el efecto de dimensionar el sistema de E/S o cambiar componentes como la red de almacenamiento, dispositivos de E/S, entre otros [9].

Simulación de arquitecturas de Cloud Computing

CloudSim [10] [11] es un Framework de simulación generalizado y extensible que permite el modelado y la simulación de diferentes infraestructuras y servicios de aplicaciones de Cloud Computing. Un ejemplo de utilización es la simulación de muchos centros de datos.

Su arquitectura consiste en entidades específicas que se representan como clases Java que pueden ser heredadas o instanciadas. Estas clases representan centros de datos, hosts físicos, máquinas virtuales, servicios a ejecutar en los centros de datos y servicios en la nube de usuarios [12] [13]. Además, CloudSim soporta la inserción dinámica de los elementos de simulación y proporciona aplicaciones de paso de mensajes y la topología de la red del centro de datos.

La versatilidad de CloudSim es la principal ventaja del sistema. La integración de nuevos parámetros y conceptos de la simulación es implementada desde abstracciones preestablecidas convenientemente por

los autores. Las abstracciones principales son SimEvent [14], SimEntity [15], DataCenterCharacteristics y Vm.

El aporte de un desarrollo de nuevos actores al componente Vm (Virtual Machine) posibilita que por medio de simulación se obtengan métricas de entrada/salida. Las estadísticas que aporta la nueva implementación de Vm dan soporte a un espacio de memoria ram en tiempos de simulación.

Simulación, Salud y Ciencia de Datos

Al analizar las necesidades presentes en los servicios de salud, se obtiene que dentro de una sala de urgencias se encuentre una gran diversidad de escenarios posibles donde cada uno de estos puede afectar a resultados sensibles como, por ejemplo, la tasa de mortalidad de personas. Para solucionar este problema se desarrolló un simulador con el objetivo de ser una herramienta capaz de recrear una gran cantidad de escenarios y así poder tomar decisiones rápidas.

El simulador desarrollado se centra en la prevalencia puntual de infecciones intrahospitalarias [16] en una sala de urgencias y cómo la afectan distintos factores relacionados con la gestión hospitalaria.

Para llevar a cabo el modelado del simulador se utilizó el paradigma de Modelado y Simulación basado en Agentes (ABMS). El framework utilizado es Repast Simphony [17], una herramienta especializada en ABMS, la cual provee una serie de ventajas respecto al modelado e implementación de agentes, así como también la coordinación concurrente de los agentes.

De esta manera, se clasificaron distintos agentes intervinientes en el ambiente de una sala de urgencias, como los pacientes y médicos, entre otros. Dichos agentes definen su comportamiento mediante máquinas de estado, las cuales determinan las acciones correspondientes tanto a la atención hospitalaria como también al estado de salud respecto de una enfermedad intrahospitalaria específica.

La propagación de la infección intrahospitalaria estará modelada mediante interacciones entre estos agentes, por ejemplo, uno de los focos de infección más importante de las salas de urgencias son las salas de espera. La forma más efectiva para calibrar el simulador se da mediante el grado de interacción de agentes, ajustando así la tasa de transmisión de la enfermedad en cuestión.

Una vez calibrado el simulador, se pueden obtener resultados. El trabajador de la salud dispone de diversos parámetros para configurar la simulación, por ejemplo, la cantidad de pacientes, la cantidad de camas disponibles, etc. Se pueden obtener resultados analizando distintos posibles cuellos de botella, configurando la cantidad de médicos clínicos disponibles, la cantidad de recepcionistas encargados de la admisión, el triage, entre otros. De esta manera, es posible la toma de decisiones respecto a la asignación de recursos y personal para agilizar la estancia hospitalaria de los pacientes y evitar los contagios producidos por sus interacciones.

Además, se analiza y trabaja la optimización y paralelización del simulador con los ambientes Repast HPC para clusters y Flame para ejecución sobre GPU, obteniendo mejor performance en términos de tiempos.

Relacionado a Ciencia de Datos, se trabaja en conjunto con el Hospital San Roque de La Plata en el análisis de datos. Se analizaron los datos obtenidos y se realizaron modificaciones o agregaciones de campos previamente calculados partiendo de la información básica, como segmentación de edad en base a la fecha de nacimiento comparado con el día del registro de ingreso. Además, es posible discriminar por por grupo etáreo y sexo, por ejemplo.

También se realizaron estudios de ingresos diarios, mensuales y anuales, para entender la periodicidad de una persona promedio en asistencias. Como el flujo promedio de personas por turnos de atención, para lo cual se podría definir la cantidad de personal por turnos y tipos de urgencias y de enfermedades relevadas.

El resultado final es un panel de control y un conjunto de datos limpio que es utilizado como entrada para el simulador previamente mencionado.

Líneas de Investigación, Desarrollo e Innovación

Temas de Estudio e Investigación

- Arquitecturas multiprocesador para procesamiento paralelo: multiprocesador de memoria compartida, multiprocesador on-chip de memoria distribuida. Multicore, Clusters, Clusters de multicore. Grid. Cloud.

- Arquitectura de E/S paralela considerando el software, hardware, comunicaciones entre módulos y dispositivos de almacenamiento.
- Nuevos aportes de desarrollos que mejoren los modelos de simulaciones con CloudSim para el análisis de la performance en sistemas de arquitecturas de software de Cloud Computing.
- Modelado y simulación para la administración de sistemas de salud.

Resultados y Objetivos

Investigación experimental

- Diseño y desarrollo de modelos mediante técnicas de Modelado y simulación basada en agentes (ABMS) para analizar el comportamiento de las distintas capas de la pila de software de E/S.
- Análisis, modelado e implementación de las operaciones típicas de E/S: read, write, open, close, flush. Se ha utilizado Amazon Web Services para creación de cluster virtuales y obtener métricas de la pila de software de E/S.
- Incorporación de tiempos de entrenamiento, obtenidos en AWS, en el simulador. Con esto se logra una salida más detallada y un método para validar tiempos y métricas del simulador con AWS.
- Implementación de comandos para ejecutar desde command center en NetLogo. Con esto se logró sintetizar el benchmark IOR correspondiente a la capa de aplicación de la pila de E/S, logrando introducir nuevos parámetros como tamaño de archivo y cantidad de nodos de E/S (metadata server y data server). De ejecutar este comando, se obtiene una nueva salida similar a la del benchmark IOR [18].
- Obtención de un método de desarrollo de nuevos actores genéricos CloudSim que mejoran el modelado y la producción de estadísticas virtuales. Implementación de la entidad InitEntity que procesa instrucciones en el espacio de memoria de las máquinas virtuales. Se vinculan exitosamente los tiempos de procesamiento de un cloudlet y las instrucciones en las máquinas virtuales [19].
- Diseño de prototipos de pruebas para estudiar entrada/salida como en el área de salud, utilizando microcontroladores [20] [21].
- En el área de tiempo real se trabajaron con microcontroladores en el reconocimiento de audio y en el control de escenarios 3D con Unity [22] [23].
- Modelado y simulación de procesos industriales, en específico sistemas de banda clasificadora [24].
- Contraste de una simulación de un cluster en la nube y uno idéntico desplegado en un sistema de cloud computing público [25].
- Modelado de infraestructuras de sistemas de salud. Implementación de un simulador para analizar el contagio de enfermedades intrahospitalarias [26] [27].
- Implementación de un panel de control con la herramienta PowerBI [28] para la toma de decisiones de un hospital a partir de un conjunto de datos real.

Formación de Recursos Humanos

Dentro de la temática de la línea de I/D se participa en el dictado de la carrera de Ingeniería en Informática de la UNAJ. También aportan trabajos de alumnos de las materias Redes de Computadoras 2 y Programación en Tiempo Real. Por otro lado, algunos integrantes participan en el dictado de la Diplomatura en Ciencia de Datos de la UNAJ.

Durante 2024 se han realizado publicaciones nacionales [29][30][31] e internacionales [18]. En esta línea de I/D existe cooperación a nivel nacional e internacional. Hay dos investigadores realizando estudios de postgrado, dos becarios de Iniciación a la Investigación UNAJ y 2 alumnos avanzados de grado colaborando en las tareas.

Referencias

1. Grama A, Gupta A, Karypis G, Kumar V. “Introduction to parallel computing”. SecondEdition. Pearson Addison Wesley, 2003.
2. R. Galeano, C. Villalba, D. Rexachs, E. Luque. Agent-Based Model to Simulate Outpatient’s Consultations at the "Hospital de Clínicas". The Eighth International Conference on Advances in System Simulation (SIMUL 2016). 1:46-51
3. H Hennessy, J. L., Patterson, and D. A., Computer Architecture, Fourth Edition: A Quantitative Approach. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.,

2006.
4. J. M. May, Parallel I/O for high performance computing. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2001.
5. V. Balaji, Earth system modelling – Volume 4. IO and Postprocessing. Springer, 2013.
6. D. Encinas, Utilización de un reloj global para el modelado de un ambiente simulado distribuido. XVIII Congreso Argentino de Ciencias de la Computación. 2012
7. D. Encinas, Simulación de una red CAN para dimensionar las comunicaciones de una IMU. VII Congreso Argentino de Tecnología Espacial. 2013.
8. D. Black, SystemC: From the Ground Up. Second Edition, Springer, 2010.
9. D. Encinas et al.,Modeling I/O System in HPC: An ABMS Approach. The Seventh International Conference on Advances in System Simulation (SIMUL), ISBN: 978-1-61208-442-8, 2015.
10. R. Calheiros, R. Ranjan, A. Beloglazov, C. De Rose and R. Buyya “CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms” Published online 24 August 2010 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/spe.995.
11. http://www.cloudbus.org/cloudsim 2018.
12. Hamza Ouarnoughi, Jalil Boukhobza, Frank Singhoff, Stephane Rubini, Erwann Kassis. “Considering I/O Processing in CloudSim for Performance and Energy Evaluation”. OpenStack Cloud Software: Open source software for building private and public clouds. © Springer International Publishing AG 2016 M. Taufer et al. (Eds.): ISC High Performance Workshops 2016, LNCS 9945, pp. 591–603, 2016. DOI: 10.1007/978-3-319-46079-6 40.
13. Parikh, Hawanna, Haleema. P.K, Jayasubalakshmi.R and N.Ch.S.N.Iyengar. School of Computing Science and engineering Vellore Institute of Technology, Tamil Nadu,“Virtual Machine Allocation Policy in Cloud Computing Using CloudSim in Java.” 2015.
14. www.icsa.inf.ed.ac.uk/research/groups/hase/simjava 2018.
15. F. Howell, R Mc Nab. A discrete event simulation library for java. International Conference on Web-Based Modeling and Simulation. 1998.
16. Prevalencia puntual de infección nosocomial, disponible en internet: https://www.researchgate.net/publication/242363760_Prevalencia_puntual_de_infeccion_nosocomial. Fecha: 29/03/2020
17. Repast Simphony Frequently Asked Questions, disponible en internet: https://repast.github.io/docs/RepastFAQ/RepastFAQ.html. Fecha: 29/03/2020
18. D. Encinas, M. Naiouf, A. D. Giusti, R. Bond, S. Mendez, D. Rexachs, and E. Luque. Agent Based Simulation as an Efficient Way for HPC I/O System Tuning. En International Conference on Computational Science. Cham: Springer Nature Switzerland, 2024. p. 129-136.
19. D. Rosatto, R. Bond, M. Belizán, M. Morales, D. Encinas. Modelado y simulación de arquitecturas de Cloud Computing con CloudSim: comunicación entre entidades. XXIII Congreso Argentino de Ciencias de la Computación. ISBN 978-950-34-1539-9. 2017
20. L. Barrera, M. Rodriguez, R. Bond, M. Morales, D. Encinas Diseño de un oxímetro de pulso. Prototipo de pruebas. XXVIII Congreso Argentino de Ciencias de la Computación. ISBN 978-987-1364-31-2. 2022
21. M. López, R. Bond, D. Encinas. Diseño de un prototipo de pruebas de Entrada/Salida con Arduino. 10º Congreso Nacional de Ingeniería Informática/Sistemas de Información (CoNaIISI 2022). Entre Ríos. 2022
22. Rodrigo Lago; Verónica Scholz; Roman Bond; Martin Morales; Diego Encinas. Experiencia de desarrollo de una aplicación de reconocimiento de canciones mediante la técnica de huella de audi". En XXIX Congreso Argentino de Ciencias de la Computación CACiC, Argentina. 2023.
23. David Verón; Roman Bond; Martin Morales; Diego Encinas. Control de iluminación en un entorno 3D con Unity. En 11° Congreso Nacional de Ingeniería Informática y Sistemas de la Información, Argentina. 2023.
24. Alan Dell Orto Marengo, Emanuel Julio, Diego Rolón, Román Bond, Martin Morales, Diego Encinas.Simulación de sistema de reciclaje mediante banda clasificadora. En 11° Congreso Nacional de Ingeniería Informática y Sistemas de la Información, Argentina. 2023.
25. T. Rosales, J. Spinelli, M. Di Nardo, R. Bond, D. Rosatto, D. Encinas, F. Romero. Análisis de una plataforma de simulación para Cloud Computing. Un caso de estudio. XXVI Congreso Argentino de Ciencias de la Computación. 2020
26. L. Maccallini, Encinas, and F. Romero. “An Approach to the Modeling and Simulation of Intra-Hospital Diseases”. Journal of computer science and technology (ISSN 1666-6038), vol. 21, num. 2, págs. 157-169, doi. 10.24215/16666038.21.e14, 2021.
27. J. Baez, A. Barreto, B. Galarza, M. Morales, D. Encinas. Simulación para estimar propagación de enfermedades. 8º Congreso Nacional de Ingeniería Informática/Sistemas de Información (CoNaIISI 2020). San Francisco, Córdoba. 2020
28. Becker & Gould (2019) Microsoft Power BI: Extending Excel to Manipulate, Analyze, and Visualize Diverse Data, Serials Review, 45:3, 184-188.
29. Albornoz, G.; Bond, R.; Morales, M.; Encinas, D. “Implementación de un Sistema de Reconocimiento Emocional Facial con Python y OpenCV”. 12º Congreso Nacional de Ingeniería Informática/Sistemas de Información (CoNaIISI 2024). Catamarca.
30. Emiliano Camer; Martin Páez; Nicolás Vitek; Román Bond; Daniel Martín Morales; Diego Encinas. “Estudio de simulación de procesos aplicado a la optimización de sistemas de producción”. En XXX Congreso Argentino de Ciencias de la Computación CACiC, Argentina. 2024.
31. Yamila Ferrufino; Román Bond; Daniel Martín Morales; Diego Encinas. “Evaluación de la Eficiencia Energética y el Impacto de las Aplicaciones de Redes Sociales”. En XXX Congreso Argentino de Ciencias de la Computación CACiC, Argentina. 2024.

Algoritmos paralelos en computación de altas prestaciones. Fundamentos, construcción y evaluación de rendimiento

Marcelo Naiouf[1], Armando De Giusti[1][2], Laura De Giusti[1][3], Franco Chichizola[1], Victoria Sanz[1][3], Adrián Pousa[1], Enzo Rucci[1][3], Mariano Sánchez[1], Manuel Costanzo[1], Emmanuel Frati[1][4], Adriana Gaudiani[5]

[1]Instituto de Investigación en Informática LIDI (III-LIDI)
Facultad de Informática – Universidad Nacional de La Plata
50 y 115, La Plata, Buenos Aires
Comisión de Investigaciones Científicas de la Pcia. de Buenos Aires (CIC)
526 e/ 10 y 11 La Plata Buenos Aires

[2] CONICET – Consejo Nacional de Investigaciones Científicas y Técnicas
[3]CIC – Comisión de Investigación Científica de la Provincia de Buenos Aires
[4] Universidad Nacional de Chilecito
[5] Universidad Nacional de General Sarmiento
{mnaiouf, degiusti, ldgiusti, francoch, vsanz, apousa, erucci, msanchez, mcostanzo, fefrati}@lidi.info.unlp.edu.ar, agaudi@ungs.edu.ar

RESUMEN

El eje central de la línea de I/D es investigar en temas de cómputo paralelo y distribuido de alto desempeño, tanto en lo referido a los fundamentos como a la construcción, evaluación y optimización de las aplicaciones en arquitecturas multiprocesador. Se aplican los conceptos en problemas numéricos y no numéricos de cómputo intensivo y/o sobre grandes volúmenes de datos con el fin de obtener soluciones de alto rendimiento.
También incluye la construcción de ambientes para la enseñanza de la programación concurrente y paralela.
En la dirección de tesis de postgrado existe colaboración con el grupo HPC4EAS (High Performance Computing for Efficient Applications and Simulation) de la Universidad Autónoma de Barcelona y el Departamento de Arquitectura de Computadores y Automática de la Universidad Complutense de Madrid, entre otros.
Palabras clave: Cómputo paralelo y distribuido de altas prestaciones. Algoritmos paralelos y distribuidos. Arquitecturas multiprocesador. Ambientes de enseñanza.

CONTEXTO

La línea de I/D que se presenta es parte del Proyecto "Computación de Alto Desempeño y Distribuida: Arquitecturas, Algoritmos, Tecnologías y Aplicaciones en HPC, Fog-Edge-Cloud, Big Data, Robótica, y Tiempo Real" del III-LIDI acreditado por el Ministerio de Educación, y de proyectos acreditados y subsidiados por la Facultad de Informática de la UNLP. Además, existe cooperación con Universidades de Argentina, Latinoamérica y Europa a través de proyectos acreditados. Asimismo, el III-LIDI forma parte del Sistema Nacional de Cómputo de Alto Desempeño (SNCAD).

1. INTRODUCCIÓN

El área de cómputo de altas prestaciones (HPC, High-Performance Computing) es clave dentro de las Ciencias de la Computación, debido al creciente interés por el desarrollo de soluciones a problemas con alta demanda computacional y de almacenamiento, produciendo transformaciones profundas en las líneas de I/D [1].
El rendimiento en este caso está relacionado con dos aspectos: las arquitecturas de soporte y los algoritmos que hacen uso de las mismas, y el desafío se centra en cómo aprovechar las prestaciones obtenidas a partir de la evolución de las arquitecturas físicas. En esta línea la mayor importancia está en los algoritmos paralelos y en los métodos utilizados para su construcción y análisis a fin de optimizarlos.
Uno de los cambios de mayor impacto ha sido el uso de manera masiva de procesadores con más de un núcleo (*multicore*), produciendo plataformas distribuidas híbridas (memoria compartida y distribuida) y generando la necesidad de desarrollar sistemas operativos, lenguajes y algoritmos que las usen adecuadamente. También creció la incorporación de placas aceleradoras a los sistemas multicore constituyendo plataformas paralelas de memoria compartida con paradigma de programación propio asociado como pueden ser las unidades de procesamiento gráfico (GPU, Graphic Processing Unit) de NVIDIA y AMD, los coprocesadores Xeon Phi de Intel [2] o los aceleradores basados en circuitos integrados reconfigurables (FPGAs, Field Programmable Gate Array) [3]. Asimismo, los entornos de computación cloud introducen un nuevo foco desde el punto de vista del HPC, brindando un soporte "a medida" sin la necesidad de adquirir el hardware.
En el último tiempo ha cobrado relevancia la investigación en el área de la computación cuántica [4][5]. La idea de una computadora cuántica surge de aplicar leyes de la física cuántica en computación. Existe mucha expectativa en este tipo de máquinas, con gran capacidad de paralelismo, ya que se espera que resuelvan algunos problemas más eficientemente que las computadoras clásicas.
La creación de algoritmos paralelos en arquitecturas multiprocesador no es un proceso directo [6]. El costo puede ser alto en términos del esfuerzo de programación y

el manejo de la concurrencia adquiere un rol central en el desarrollo. Si bien en las primeras etapas el diseñador de una aplicación paralela puede abstraerse de la máquina sobre la que ejecutará el algoritmo, para obtener buen rendimiento debe tenerse en cuenta la plataforma de destino. En las máquinas multiprocesador, se deben identificar las capacidades de procesamiento, interconexión, sincronización y escalabilidad. La caracterización y estudio de rendimiento del sistema de comunicaciones es de interés para la predicción y optimización de performance, así como la homogeneidad o heterogeneidad de los procesadores [7].

Muchos problemas algorítmicos se vieron impactados por los multicore y clusters de multicore. El desarrollo de algoritmos que aprovechen adecuadamente las arquitecturas motiva el estudio de performance en sistemas híbridos. Es necesario estudiar la utilización de lenguajes y bibliotecas ya que aún no se cuenta con un standard, aunque puede mencionarse el uso de los tradicionales MPI, OpenMP y Pthreads [8][9] o los más recientemente explorados UPC, Chapel y Titanium del modelo PGAS [10].

La combinación de arquitecturas de múltiples núcleos con aceleradores dio lugar a plataformas híbridas con diferentes características. Actualmente, se plantea la idea de incorporar computadoras cuánticas a estas plataformas híbridas. Varios autores suponen que las computadoras cuánticas no reemplazarán a las computadoras clásicas, sino que se incorporarán como aceleradores a sistemas paralelos existentes [11]. Más allá del acelerador utilizado, la programación de estas plataformas representa un desafío. Para lograr aplicaciones de alto rendimiento, los programadores enfrentan dificultades como: estudiar características específicas de cada arquitectura y aplicar técnicas de programación y optimización particulares de cada una, lograr un balance de carga adecuado entre los dispositivos de procesamiento y afrontar la ausencia de estándares para este tipo de sistemas.

Por otra parte, los avances en las tecnologías de virtualización han llevado a que Cloud Computing sea una alternativa a los tradicionales sistemas de cluster [12]. El uso de cloud para HPC presenta desafíos atractivos, brindando un entorno reconfigurable dinámicamente sin la necesidad de adquirir hardware, y es una excelente plataforma para testear escalabilidad de algoritmos.

Métricas de evaluación del rendimiento y balance de carga

La diversidad de opciones vuelve complejo el análisis de performance de los Sistemas Paralelos, ya que los ejes sobre los cuales pueden compararse dos sistemas son varios. Existe un gran número de métricas para evaluar el rendimiento, siendo las tradicionales: tiempo de ejecución, speedup, eficiencia. Por su parte, la *escalabilidad* permite capturar características de un algoritmo paralelo y la arquitectura en que se lo implementa. Posibilita testear la performance de un programa sobre pocos procesadores y predecirla en un número mayor, así como caracterizar la cantidad de paralelismo inherente en un algoritmo.

Un aspecto de interés que se ha sumado como métrica, a partir de las plataformas con gran cantidad de procesadores, es el del *consumo* y la *eficiencia energética* [13]. Muchos esfuerzos están orientados a tratar el consumo como eje de I/D, como métrica de evaluación, y también a la necesidad de metodologías para medirlo.

El objetivo principal del cómputo paralelo es reducir el tiempo de ejecución haciendo uso eficiente de los recursos. El *balance de carga* es un aspecto central y consiste en, dado un conjunto de tareas que comprenden un algoritmo y un conjunto de procesadores, encontrar el mapeo (asignación) de tareas a procesadores tal que cada una tenga una cantidad de trabajo que demande aproximadamente el mismo tiempo, y esto es más complejo si hay heterogeneidad. Dado que el problema general de mapping es *NP*-completo, pueden usarse enfoques que dan soluciones subóptimas aceptables. Las técnicas de planificación a nivel micro (dentro de cada procesador) y macro (en un cluster) deben ser capaces de obtener buen balance de carga. Existen técnicas estáticas y dinámicas cuyo uso depende del conocimiento que se tenga sobre las tareas de la aplicación.

2. LÍNEAS DE INVESTIGACIÓN, DESARROLLO E INNOVACIÓN

- Investigar en temas de cómputo paralelo y distribuido de alto desempeño, en lo referido a los fundamentos y a la construcción y evaluación de las aplicaciones. Esto incluye los problemas de software asociados con el uso de arquitecturas multiprocesador:
  - Lenguajes, modelos y paradigmas de programación paralela (puros e híbridos a distintos niveles).
  - Asignación de procesos a procesadores optimizando el balance de la carga de procesamiento.
  - Métricas de evaluación de complejidad y rendimiento: speedup, eficiencia, escalabilidad, consumo energético, costo de programación.
- Construir, evaluar y optimizar soluciones utilizando algoritmos concurrentes, paralelos y distribuidos sobre diferentes plataformas de software y arquitecturas con múltiples procesadores:
  - Arquitecturas de trabajo homogéneas, heterogéneas e híbridas: multicores, clusters, GPU, Xeon Phi, FPGA, entornos cloud y nuevas plataformas como las computadoras cuánticas.
  - Aplicar los conceptos en problemas numéricos y no numéricos de cómputo intensivo y/o sobre grandes volúmenes de datos (aplicaciones científicas, búsquedas, simulaciones, imágenes, realidad virtual y aumentada, bioinformática, big data, n-body).
- Analizar y desarrollar ambientes para la enseñanza de programación concurrente y paralela.

- Caracterizar diferentes modelos de arquitecturas paralelas.
- Representar distintos modelos de comunicación/sincronización.
- Definir métricas de evaluación de rendimiento y eficiencia energética.

3. RESULTADOS OBTENIDOS/ESPERADOS

- Desarrollar y optimizar algoritmos paralelos sobre diferentes modelos de arquitectura. En particular, en aplicaciones numéricas y no numéricas de cómputo intensivo y tratamiento de grandes volúmenes de datos.
- Estudiar y comparar los lenguajes sobre las plataformas multiprocesador para diferentes modelos de interacción entre procesos.
- Investigar la paralelización en plataformas que combinan clusters, multicore y aceleradores. Comparar estrategias de distribución de trabajo teniendo en cuenta las diferencias en potencias de cómputo y comunicación, dependencia de datos y memoria requerida.
- Evaluar la performance (speedup, eficiencia, escalabilidad, consumo energético) de las soluciones propuestas. Analizar el rendimiento de soluciones paralelas a problemas con diferentes características (dependencia de datos, relación cómputo / comunicación, memoria requerida).
- Mejorar y adecuar las técnicas disponibles para el balance de carga (estático y dinámico) entre procesos a las arquitecturas consideradas.

En este marco, pueden mencionarse los siguientes resultados:

- Para la experimentación se han utilizado y analizado diferentes arquitecturas homogéneas o heterogéneas, incluyendo multicores, cluster de multicores (con 128 núcleos), GPU y cluster de GPU, Xeon Phi y FPGA.
- Se experimentó la paralelización en arquitecturas híbridas, con el objetivo de estudiar el impacto del mapeo de datos y procesos, así como de los lenguajes y librerías.
- Respecto de las aplicaciones y temas estudiados, se trabajó fundamentalmente con los siguientes problemas:

➢ **Aceleración de aplicaciones con cómputo colaborativo CPU-GPU.** Las computadoras comerciales actuales incluyen decenas de cores y al menos una GPU. El uso de ambas unidades de procesamiento de forma colaborativa puede mejorar significativamente el rendimiento de una aplicación. Sin embargo, esto supone un desafío para los programadores ya que dichas unidades difieren en arquitectura, modelo de programación y rendimiento. En [14] se propuso un modelo híbrido para estructurar código a ser ejecutado sobre un sistema heterogéneo con múltiples cores y 1 GPU (utilizando todos los recursos disponibles). Utilizando este modelo se desarrolló un algoritmo paralelo de pattern matching para sistemas heterogéneos CPU-GPU [15]. Los resultados revelaron que este algoritmo supera en rendimiento a trabajos previos, desarrollados para sistemas multicore y GPUs, para datos de tamaño considerable. En [16] se presentó una solución al problema de pattern matching que aprovecha toda la potencia computacional de los procesadores Intel Xeon Phi KNL 7230 mediante el uso de SIMD y paralelismo de hilos. Se mostró que el algoritmo propuesto alcanza aceleraciones significativas. En trabajos futuros interesa investigar sobre el cómputo colaborativo incluyendo este tipo de arquitecturas.

➢ **Bioinformática y biología computacional.** Estas áreas incluyen aplicaciones que se nutren de las capacidades de HPC para brindar soluciones en un tiempo de respuesta aceptable, siendo cada día más debido al crecimiento exponencial de datos genómicos en los últimos años. Es posible encontrar aplicaciones en alineamiento de secuencias, acoplamiento y dinámica molecular, predicción y búsqueda de estructuras moleculares, entre otras.

➢ **Alineamiento de secuencias.** Si bien el algoritmo de Smith-Waterman es considerado el método de alineamiento más preciso, este algoritmo resulta costoso debido a su complejidad computacional cuadrática. Dado el surgimiento del estándar SYCL y de implementaciones como Intel oneAPI, se migró una versión CUDA de este algoritmo a su equivalente en SYCL, analizando el esfuerzo de programación requerido y la relación costo-beneficio entre overhead y portabilidad para diferentes arquitecturas CPU y GPUs [17][18][19]. A futuro, se propone extender el estudio realizado contemplando otras generaciones de GPUs y arquitecturas como FPGAs.

➢ **Identificación de biomarcadores.** Entre todas las plataformas, servicios y paquetes de software disponibles para esta aplicación, se encuentra Multiomix [20], quien se destaca del resto por facilitar el proceso de integración y minería de datos oncogenómicos públicos y cargados por los usuarios, proporcionando una interfaz gráfica de usuario amigable para los usuarios no expertos. A pesar de ello, Mutiomix aun funciona de forma secuencial por ser un desarrollo reciente (una ejecución habitual en Multiomix puede tomar varias horas para datasets de pequeño porte). Se propone acelerar el procesamiento de Multiomix de manera de disminuir los tiempos de respuesta y así mejorar la productividad de sus usuarios.

➢ **Cálculo de los caminos mínimos**. Es uno de los problemas básicos y de mayor antigüedad de la teoría de grafos teniendo aplicación en el dominio de las comunicaciones, del ruteo de tráfico, de la bioinformática, entre otros. El algoritmo de Floyd-Warshall (FW) permite computar la distancia mínima entre todos los pares de un grafo. Además de poseer una

alta demanda de ancho de banda, FW resulta costoso computacionalmente al ser $O(n^3)$. Habiendo desarrollado una implementación optimizada para arquitecturas x86 [21], se diseñó y desarrollo una librería C que permite usar este código de manera sencilla y práctica en aplicaciones de terceros, dando soporte a diferentes tipos de datos y formatos de archivos de entrada y salida [22].

- **Problemas de optimización de simulación de sistemas dinámicos complejos mediante heurísticas**. En una primera etapa se trabajó con una metodología de sintonización de modelos físicos usando para las experimentaciones un modelo de cauce de ríos. Estos modelos requieren de técnicas que detecten la falta de calibración de sus parámetros debido a muchas clases de incertidumbre que impactan en la calidad de los resultados. Se buscó un conjunto ajustado de parámetros que puedan sintonizar el simulador y se propuso un metamodelo de optimización mediante simulación basado en una heurística combinada entre un método de búsqueda Monte Carlo y un algoritmo de agrupamiento KMeans. Se lograron resultados alentadores que mejoraron la calidad de la simulación con gran reducción de los recursos y tiempo de cómputo. Se actuó "por proximidad" entre las distintas configuraciones de parámetros, lo cual fue incluido en el método combinado. En una segunda etapa se actuó por "similitud" de los valores de los parámetros, aprovechando que los valores de magnitudes físicas en regiones contiguas del dominio tienen poca variación. Se agrega conocimiento del sistema simulado, lo que disminuyó considerablemente el espacio de búsqueda de los parámetros y el costo computacional. Actualmente, se aumentó este conocimiento del sistema aprovechando el conocimiento de eventos disruptivos del pasado y su sintonización construyendo una base histórica de eventos que pueden aprovecharse en eventos actuales, disminuyendo aún más el costo computacional de calibración del modelo [23][24].

- **Ambientes para la enseñanza de concurrencia.** Se desarrolló el entorno CMRE para la enseñanza de programación concurrente y paralela a partir de cursos iniciales en carreras de Informática. Incluye un entorno visual que representa una ciudad en la que pueden definirse varios robots que interactúan. Combina aspectos de memoria compartida y distribuida mediante instrucciones para bloquear y liberar esquinas de la ciudad y el concepto de pasaje de mensajes a través de primitivas de envío y recepción. Además, se incluyen los conceptos de heterogeneidad (diferentes velocidades de los robots) y consumo energético [25]. Se ha integrado con el uso de robots físicos (Lego Mindstorm 3.0) que ejecutan en tiempo real las mismas instrucciones que los robots virtuales y se comunican con el entorno mediante bluetooth [26]. Se ha ampliado para incorporar conceptos básicos de computación en la nube (Cloud Computing) [26]. Actualmente, se está desarrollando una nueva herramienta para la enseñanza de programación concurrente en cursos avanzados. Su objetivo principal es visualizar los conceptos de sincronización y comunicación entre procesos.

- **Cifrado de grandes volúmenes de datos.** Hoy en día, la cantidad de datos sensibles que se generan para ser almacenados y/o transmitidos a través de la red aumenta constantemente. Para proteger los datos confidenciales de amenazas potenciales, se utilizan estrategias de encriptación. Además, el tiempo involucrado en el cifrado de datos está directamente relacionado con la cantidad de datos que se cifrarán y puede ser significativo. Para reducir el tiempo de cifrado es natural recurrir a soluciones de encriptación paralelas, que explotan todo el poder computacional proporcionado por las arquitecturas emergentes. AES (Advanced Encryption Standard) es uno de los algoritmos de cifrado más utilizados y el gobierno de los Estados Unidos lo considera lo suficientemente seguro como para proteger la información nacional. Hay varias implementaciones de AES, tanto en hardware como en software. En [27] se presenta una comparación de rendimiento de una solución AES basada en hardware para CPU multinúcleo con el de otras dos soluciones AES basadas en software para CPU multinúcleo y GPU, respectivamente. El primero se implementa con las nuevas instrucciones de Intel AES y el segundo con la biblioteca OpenSSL. Los resultados revelan que utilizar cómputo paralelo para el proceso de cifrado reduce significativamente el tiempo de ejecución respecto a una solución secuencial. Los resultados también muestran que el mayor rendimiento se alcanza utilizando una solución multicore con AES basado en hardware. Sin embargo, la solución por software utilizando 2 GPUs puede ser una alternativa competitiva a la solución multicore por hardware cuando se tienen pocos núcleos o una CPU que no soporta AES. Asimismo, en [28] se muestra que una solución colaborativa, que utiliza al mismo tiempo múltiples cores y 2 GPUs para el cifrado de grandes volúmenes de datos, permite alcanzar mejoras en el rendimiento.

- **Compresión de grandes volúmenes de datos.** Grandes volúmenes de datos requieren grandes volúmenes de almacenamiento. Por esta razón, se hace necesario desarrollar algoritmos de compresión que ayuden a reducir el costo de almacenamiento sin consumir demasiado tiempo. Un ejemplo de esto es el almacenamiento de datos genómicos. La genómica ha ganado relevancia ya que permite prevenir, diagnosticar y tratar enfermedades de forma personalizada. Es a partir de esta mayor demanda que se requiere una gran capacidad de almacenamiento de datos genómicos. En [29] se propuso una mejora a un algoritmo multihilo ya existente obteniendo mejoras en el tiempo de ejecución mientras se conserva un bajo ratio de compresión.

- **Computación cuántica.** Aunque las capacidades de las computadoras cuánticas actuales son muy limitadas, existen muchas líneas de investigación en el área. Estas abarcan el estudio de la fabricación de arquitecturas cuánticas eficientes, simuladores, control de errores por decoherencia, traspilación, el desarrollo de algoritmos cuánticos, la migración de algoritmos clásicos a cuánticos, integración de sistemas cuánticos y clásicos, entornos de desarrollo y lenguajes de programación cuánticos, entre otras.

Organización de Eventos

En 2024 se organizaron las XII Jornadas de Cloud Computing, Big Data & Emerging Topics (JCC-BD&ET 2024), con participación de especialistas académicos del país y el exterior, además de empresas con experiencia en Cloud Computing [30][31]. En junio de 2025 se tiene prevista la organización de las XIII JCC-BD&ET [32].

4. FORMACIÓN DE RECURSOS HUMANOS

Dentro de la temática de la línea de I/D se concluyó 1 tesis doctoral,y 2 Tesinas de Licenciatura. Se encuentran en curso en el marco del proyecto 3 tesis doctorales, 2 de maestría, y 2 Tesinas.

Se participa en el dictado de las carreras de Doctorado en Cs. Informáticas y Maestría y Especialización en Cómputo de Altas Prestaciones de la Facultad de Informática UNLP, por lo que potencialmente pueden generarse más Tesis y Trabajos Finales.

Asimismo, en grado, los integrantes de la línea de I/D dictan materias relacionadas como Programación Concurrente, Sistemas Paralelos y Taller de Programación sobre GPU.

Existe cooperación con grupos de otras Universidades del país y del exterior, y tesistas de diferentes Universidades realizan su trabajo con el equipo del proyecto.

5. BIBLIOGRAFÍA

[1]. Giles MB, Reguly I. “Trends in high-performance computing for engineering calculations”. Phil.Trans.R.Soc.A 372: 20130319. 2014. http://dx.doi.org/10.1098/rsta.2013.0319

[2]. Jeffers J., Reinders J. “Intel Xeon Phi Coprocessor High Performance Programming”. Morgan Kaufmann.

[3]. Sean Settle. “High-performance Dynamic Programming on FPGAs with OpenCL”. IEEE High Performance Extreme Computing Conf (HPEC ’13). 2013. https://doi.org/10.1016/j.physa.2017.04.159.

[4]. IBM Quantum Computing https://www.ibm.com/quantum

[5]. Google Quantum AI https://quantumai.google/quantumcomputer

[6]. McCool M. “Structured Parallel Programming: Patterns for Efficient Computation”. Morgan Kaufmann.

[7]. De Giusti L, Naiouf M., Chichizola F., Luque E., De Giusti A. “Dynamic Scheduling in Heterogeneous Multiprocessor Architectures. Efficiency Analysis”. Computer Science and Technology Series – XV Argentine Congress of Computer Science Selected Papers. La Plata (Buenos Aires): Editorial de la Universidad de La Plata (edulp). 2010. p85 - 95. isbn 978-950-34-0684-7.

[8]. Chapman B., Jost G., Van der Pas. “Using OpenMP – Portable Shared Memory Parallel Programming”. 2008. UK: MIT Press.

[9]. Hager G., Wellein G. “Introduction to HPC for Scientists and Engineers”. 2011. EEUU: CRC Press.

[10]. De Wael M., Marr S., De Fraine B., Van Cutsem T., De Meuter W. “Partitioned Global Addess Space Languages”. ACM Computing Surveys 47 (4), 2015.

[11]. Pousa, A., Sanz, V. M., Naiouf, M., & De Giusti, A. (2024). Trends in High Performance Computing and Quantum Computing. Journal of Computer Science and Technology, 24(2), e11. https://doi.org/10.24215/16666038.24.e11

[12]. Amazon Elastic Compute Cloud (Amazon EC2). http://aws.amazon.com/es/ec2/. Febrero 2013.

[13]. Balladini J., Rucci E., De Giusti A., Naiouf M., Suppi R., Rexachs D., Luque E. “Power Characterisation of Shared-Memory HPC Systems”. Computer Science & Technology Series – XVIII Argentine Congress of Computer Science Selected Papers. ISBN 978-987-1985-20-3. Pp. 53-65. EDULP, La Plata (Argentina), 2013

[14]. Sanz V., Pousa A., Naiouf M., De Giusti A. “Accelerating Pattern Matching with CPU-GPU Collaborative Computing”. Proceedings of the International Conference on Algorithms and Architectures for Parallel Processing (ICA3PP 2018), ISBN: 978-3-030-05051-1, págs. 310-322, doi. https://doi.org/10.1007/978-3-030-05051-1_22, 2018.

[15]. Sanz V., Pousa A., Naiouf M., De Giusti A. “Efficient Pattern Matching on CPU-GPU Heterogeneous Systems”. In: Wen S., Zomaya A., Yang L. (eds) Algorithms and Architectures for Parallel Processing. ICA3PP 2019. Lecture Notes in Computer Science, vol 11944. Springer, Cham. 2020.

[16]. Sanz V., Pousa A., Naiouf M., De Giusti A. “Accelerating Pattern Matching on Intel Xeon Phi Processors”. In: Qiu M. (eds) Algorithms and Architectures for Parallel Processing. ICA3PP 2020. Lecture Notes in Computer Science, vol 12452. Springer, Cham. https://doi.org/10.1007/978-3-030-60245-1_18.

[17]. Costanzo M., Rucci E., García-Sánchez C., Naiouf M., Prieto-Matías M.. “Migrating CUDA to oneAPI: A Smith-Waterman Case Study.” International Work-Conference on Bioinformatics and Biomedical Engineering 2022, ISBN 978-3-031-07802-6, págs. 103-116, Cham, doi: 10.1007/978-3-031-07802-6_9. 2022.

[18]. Costanzo M., Rucci E., García-Sánchez C., Naiouf M., Prieto-Matías M. “Comparing Performance and Portability Between CUDA and SYCL for Protein Database Search on NVIDIA, AMD, and Intel GPUs.”

2023 IEEE 35th International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD). ISBN 979-8-3503-0548-7. Págs. 141-148. Porto Alegre, Brasil. doi: 10.1109/SBAC-PAD59825.2023.00023. Octubre 2023.
[19]. Costanzo M., Rucci E., García-Sánchez C., Naiouf M., Prieto-Matías M. "Assessing opportunities of SYCL for biological sequence alignment on GPU-based systems." J Supercomput (2024). https://doi.org/10.1007/s11227-024-05907-2
[20]. Camele G., Menazzi S., Chanfreau H., Marraco A., Hasperué W., Butti M. D., Abba M.C. "Multiomix: a cloud-based platform to infer cancer genomic and epigenomic events associated with gene expression modulation". Bioinformatics, Volume 38, Issue 3, 1 February 2022, Pages 866–868, https://doi.org/10.1093/bioinformatics/btab678
[21]. S. Calderón, E. Rucci, y F. Chichizola, «Enhanced OpenMP Algorithm to Compute All-Pairs Shortest Path on X86 Architectures», en Computer Science – CACIC 2023, P. Pesado, W. Panessi, y J. M. Fernández, Eds., Cham: Springer Nature Switzerland, 2024, pp. 46-61. doi: 10.1007/978-3-031-62245-8_4.
[22]. Jerónimo Lambre, Enzo Rucci. "Librería C para Cómputo Paralelo de Caminos Mínimos en Grafos sobre Arquitecturas Multicore". XXIX Congreso Argentino de Ciencias de la Computación – CACIC 2024. Octubre de 2024. En prensa
[23]. Trigila, M., Gaudiani, A., Wong, A., Rexachs, D., Luque, E. (2023). Reduction of the Computational Cost of Tuning Methodology of a Simulator of a Physical System. In: Mikyška, J., de Mulatier, C., Paszynski, M., Krzhizhanovskaya, V.V., Dongarra, J.J., Sloot, P.M. (eds) Computational Science – ICCS 2023. ICCS 2023. Lecture Notes in Computer Science, vol 10475. Springer, Cham. https://doi.org/10.1007/978-3-031-36024-4_49.
[24]. Gaudiani, A., Wong, A., Luque, E. et al. A computational methodology applied to optimize the performance of a river model under uncertainty conditions. J Supercomput 79, 4737–4759 (2023). https://doi.org/10.1007/s11227-022-04816-6
[25]. Castro J., De Giusti L., Gorga G., Sánchez M., Naiouf M. "ECMRE: Extended Concurrent Multi Robot Environment". Computer Science – CACIC 2017. Communications in Computer and Information Science, vol 790, ISBN: 978-3-319-75213-6 978-3-319-75214-3, Springer, Cham, págs. 285-294. 2018.
[26]. De Giusti L., Chichizola F., Rodríguez Eguren S., Sánchez M., Paniego J. M., De Giusti A. "Introduciendo conceptos de Cloud Computing utilizando el entorno CMRE". Proceedings del XXII Congreso Argentino de Ciencias de la Computación (CACIC 2016) – Workshop de Innovación en Educación en Informática. 2016.
[27]. Sanz V., Pousa A., Naiouf M., De Giusti A. "Comparison of Hardware and Software Implementations of AES on Shared-Memory Architectures". In: Naiouf M., Rucci E., Chichizola F., De Giusti L. (eds) Cloud Computing, Big Data & Emerging Topics. JCC-BD&ET 2021. Communications in Computer and Information Science, vol 1444. Springer, Cham. 2021. https://doi.org/10.1007/978-3-030-84825-5_5.
[28]. Sanz, V., Pousa, A., Naiouf, M., De Giusti, A. "Performance Analysis of AES on CPU-GPU Heterogeneous Systems". In: Rucci, E., Naiouf, M., Chichizola, F., De Giusti, L., De Giusti, A. (eds) Cloud Computing, Big Data & Emerging Topics. JCC-BD&ET 2022. Communications in Computer and Information Science, vol 1634. Springer, Cham. 2022. https://doi.org/10.1007/978-3-031-14599-5_3
[29]. Sanz V., Pousa A., Naiouf M., De Giusti A. "Fast Genomic Data Compression on Multicore Machine" XII Jornadas de Cloud Computing, Big Data & Emerging Topics. DOIhttps://doi.org/10.1007/978-3-031-70807-7_1
[30]. Naiouf M., De Giusti L., Chichizola F., Libutti L. "Cloud Computing, Big Data & Emerging Topics: 12th Conference, JCC-BD&ET 2024, La Plata, Argentina, June 25-27, 2024, Revised Selected Papers". Communications in Computer and Information Science. Springer Cham. 2024. https://doi.org/10.1007/978-3-031-70807-7.
[31]. De Giusti A., Naiouf M., Chichizola F., De Giusti L. C., Libutti L. A. (eds). "Short papers of the 12th Conference on Cloud Computing, Big Data & Emerging Topics (JCC-BD&ET 2024)". 2024. https://doi.org/10.35537/10915/171300.
[32]. XIII Jornadas de Cloud Computing, Big Data & Emerging Topics. https://jcc.info.unlp.edu.ar/

Arquitecturas en la Nube: Diseño, Simulación y Despliegue.

Diego Encinas[1,2], Brian Galarza[1], Román Bond[1], Gonzalo Zaccardi[1], Nicolás Benquerença Mendes[1], Rodrigo Lago[1], Jorge Osio[1], David Duarte[1], Martin Morales[1,3]

[1]Proyecto de Investigación SimHPC - Programa TICAPPS - Instituto de Ingeniería y Agronomía - UNAJ

[2]Instituto de Investigación en Informática LIDI (III-LIDI) - Facultad de Informática - UNLP – Centro Asociado CIC

[3]Centro CodApli - Facultad Regional La Plata - UTN

dencinas@unaj.edu.ar, bgalarza@unaj.edu.ar, rbond@unaj.edu.ar, gzaccardi@unaj.edu.ar, nbenquerenca@unaj.edu.ar, rodrigazo98@gmail.com, josio@unaj.edu.ar, davito.duarte.22@gmail.com, martin.morales@unaj.edu.ar

Resumen

El objetivo de esta línea de investigación es el estudio del rendimiento de las arquitecturas tipo cloud a través del despliegue de IaaS y utilización de IaaS públicos, en particular en el área de cómputo paralelo de altas prestaciones (HPC). Enfocando en la obtención de herramientas que permitan predecir la eficiencia del sistema ante posibles escenarios. Analizando los diferentes componentes del sistema que pueden influir en las prestaciones significativamente, especialmente la entrada/salida y las comunicaciones.

Palabras clave: *Cloud Computing. OpenStack. Sistemas de Archivos en clústers. Redes definidas por Software.*

Contexto

Se presenta una línea de Investigación que es parte del Proyecto de Investigación "Simulación, Computación de Altas Prestaciones (HPC) y optimización de aplicaciones sociales – SimHPC" de la Universidad Nacional Arturo Jauretche (UNAJ), acreditado por resolución interna 183/21. Además, el proyecto colabora con el Programa "Tecnologías de la información y la comunicación (TIC) en aplicaciones de interés social" – TICAPPS de la UNAJ.

Introducción

Cloud Computing es un paradigma que ha estado en constante crecimiento. Cada vez más compañías y grupos de investigación trabajan en conjunto con el fin de explotar las oportunidades ofrecidas por el mismo [1]. Dicho paradigma ofrece muchas ventajas, tales como el bajo costo de implementación, ya que no se necesitan computadoras de última tecnología debido a que éstas trabajan conjuntamente (Clustering) con la posibilidad de escalar horizontalmente de manera sencilla. Además, hay software Open Source disponible para los nodos en el clúster como las infraestructuras Eucalyptus, OpenNebula, CloudStack u OpenStack integradas con GNU/Linux y compatibles, por ejemplo, con Amazon WebServices.

Despliegue de IaaS

Las comunicaciones en Cloud Computing son una parte fundamental del paradigma que consisten en utilizar distintos nodos y lograr hacerlos funcionar conjuntamente.

Para lograr una comunicación sincronizada entre estos nodos se propone utilizar OpenStack [2] como también OpenNebula [3].

OpenStack es una plataforma de tecnología open source que utiliza recursos virtuales agrupados para diseñar y gestionar nubes privadas y públicas a través de múltiples servicios que, de manera coordinada, cumplen diferentes propósitos para lograr el correcto funcionamiento de, por ejemplo, una ”Infraestructure as a service” (IaaS). Algunos de los servicios ofrecidos por OpenStack son: hypervisor (Nova), autentificación (Keystone), Imágenes (Glance), Dashboard (Horizon), Networking (Neutron) y block storage (Cinder). Según las necesidades se pueden requerir de ciertos servicios u otros. La Arquitectura básicamente consiste en dos tipos de nodos: “Compute Node” y “Controller Node”. Se llaman Compute Node a todos aquellos que se encargan del procesamiento de servicios específicos mientras que Controller Node es aquel que comunica a cada uno de los anteriores [4] [5] [6].

Fuel es una herramienta open source desarrollada por Mirantis en la cual se ejecuta un script que permite configurar, de manera más amigable respecto a OpenStack, los recursos que se desean otorgar a la infraestructura, como la cantidad de nodos, los núcleos de procesador, la memoria RAM, entre otros [7].

Fuel trabaja con un nodo master el cual es el encargado de controlar a los nodos slaves que contendrán la infraestructura OpenStack. Es decir, desde el nodo Fuel Master se indican qué paquetes se van a instalar en cada nodo slave (Glance, Nova-Compute, Keystone, etc.) para luego en los slaves tener armados los nodos compute y controller, sin necesidad de realizar configuraciones manuales en cada uno de los mismos.

OpenNebula es un software de código abierto que permite el despliegue de IaaS. Busca reducir la complejidad generada por OpenStack y ofrece soporte con hypervisores tales como KVM y VMware vCenter.

OpenNebula clasifica a los nodos en dos tipos, Front – end los cuales entran en contacto con los usuarios y a su vez se comunican con los nodos de la infraestructura en los cuales se lanzarán las instancias y los nodos virtualizados los que a su vez deben contar con los paquetes correspondientes de storage, autenticación y networking para poder funcionar correctamente.

La implementación de estas infraestructuras ofrece ventajas en las cuales los clústeres virtualizados trabajan en conjunto ofreciendo un buen rendimiento a bajos costos y con posibilidad de escalabilidad al poder agregar mayor cantidad de nodos para procesamiento de manera sencilla.

Sistemas de Archivos Paralelos en clústeres

Amazon Webservices [8], mediante el servicio EC2 (Amazon Elastic Compute Cloud) permite desplegar clústeres virtuales mediante instancias de VMs y almacenamiento para las mismas. Mediante este servicio y el correspondiente clúster conformado, se utiliza un sistema de archivos paralelo (PVFS2) que permite la gestión de datos particionados y distribuidos en los distintos nodos, mediante múltiples tareas de una aplicación ejecutada sobre el clúster.

PVFS2 utiliza una estructura cliente-servidor. Dependiendo de su rol, existen tres tipos de nodos dentro de un clúster con este sistema de archivos: servidores de datos, servidores de metadatos y clientes, en donde cada uno de los nodos puede cumplir los tres roles.

PVFS2 [9] contiene una herramienta interna que permite volcar información en logs durante la ejecución de tareas dentro del sistema de archivos. Por cada uno de los nodos, el administrador del clúster puede obtener información de depuración (GOSSIP) conformado por registros de debug, de acceso, contadores de rendimiento y errores producidos durante la ejecución. Referido a los contadores de rendimiento, se obtienen mediante el software de monitoreo Atop. Esta herramienta permite obtener reportes de la actividad de los procesos y la utilización de los diferentes recursos del sistema (memoria, disco, CPU, red, etc) [10].

Asimismo, es posible especificar los datos a obtener en función de las distintas capas de PVFS2, del rol específico de cada nodo (cliente-servidor), así como también de otros factores relacionados con el funcionamiento interno y operaciones asociadas a la gestión de archivos.

Teniendo conocimiento del funcionamiento del código fuente correspondiente a PVFS2 y del modo en el que realiza la escritura de registros en los logs, resulta factible realizar modificaciones que permitan obtener otro tipo de parámetros de interés a partir de esta herramienta.

Finalmente se propone el análisis, uso y configuración de distintas herramientas no invasivas para determinar la performance del sistema de archivos en clústeres virtuales.

Redes definidas por Software

Las redes definidas por software (SDN) son un paradigma de gestión y administración de redes por medio de software que permiten tener un control más flexible con respecto al control del tráfico de datos por medio de hardware ya que permite cambiar en tiempo real las normas y políticas establecidas en la red[11].

Mininet[12] es un emulador de redes SDN open source que permite generar tráfico artificial entre nodos virtuales de la red.

Se ha llevado a cabo el montaje de un laboratorio SDN utilizando Mininet junto con Amazon WebServices con el fin de analizar el comportamiento de este nuevo paradigma de redes de computadoras en un entorno de Cloud Computing.[13]

Líneas de Investigación, Desarrollo e Innovación

Temas de Estudio e Investigación

- Arquitecturas multiprocesador para procesamiento paralelo: multiprocesador de memoria compartida, multiprocesador on-chip de memoria distribuida. Multicore, Clusters, Clusters de multicore. Grid. Cloud.
- Plataformas de software para implementar y administrar Clouds públicos, privados e híbridos.
- Sistemas de Archivos Paralelos.

Resultados y Objetivos

Investigación experimental

- Implementación de un IaaS encargado de realizar operaciones en procesamiento paralelo aumentando la eficiencia y reduciendo los costes generados.

- Implementación de OpenStack Dashboard y de un sistema desarrollado para poder controlar/administrar de manera visual (web) y más básica cada uno de los servicios.
- Implementación de OpenNebula en un sistema con las mismas características que el implementado por OpenStack con el fin de poder realizar pruebas en entornos similares.
- Utilización de Fuel para administrar OpenStack como sistema de administración de nube (Cloud Computing) a partir de la infraestructura de 2 nodos compute y e1 controller [14].
- Análisis del rendimiento de un Cloud privado en la ejecución de instancias personalizadas.
- Ejecutar diferentes benchmarks en la infraestructura desplegada sobre OpenStack y OpenNebula para así comparar los resultados obtenidos de ambas infraestructuras y poder realizar un análisis del rendimiento en cada caso.
- Introducción al estudio e integración de sensores físicos y los servicios en la nube [15].
- Medición de servicios en la nube enfocados a IaaS y PaaS [16].
- Análisis y configuración de clústeres virtuales. Análisis y configuración de herramientas no invasivas para la obtención de métricas en las distintas capas de software de los sistemas de archivos paralelos.
- Utilización de otros sistemas de archivos paralelos como Lustre [17] y Beegfs [18] para obtener métricas en Metadataservidores [19] [20] [21].
- En el área de tiempo real se trabajó en el reconocimiento de audio mediante la técnica huella de audio [22]. Por otro lado, se estudió mediante simulación el funcionamiento de una banda clasificadora utilizada en reciclaje [23].

Formación de Recursos Humanos

Dentro de la temática de la línea de I/D se participa en el dictado de la carrera de Ingeniería Informática de la UNAJ. También aportan trabajos de alumnos de las materias Sistemas Operativos 1, Redes de Computadoras 2, Programación en Tiempo Real y Organización y Arquitecturas de Computadoras. Por otro lado, algunos integrantes participan en el dictado de la Maestría y Diplomatura en Ciencia de Datos de la UNAJ.

Durante 2024 se han realizado publicaciones nacionales [24][25][26]. En esta línea de I/D existe cooperación a nivel nacional. Hay 4 investigadores realizando carreras de postgrado y alumnos avanzados de grado colaborando en las tareas.

Referencias

1. Kondo, D., Javadi, B., Malecot, P., Cappello, F., Anderson, D. P.: “Cost-benefit analysis of Cloud Computing versus desktop grids”. In: IPDPS '09 Proceedings. IEEE International Symposium on Parallel and Distributed Processing. Washington, USA (2009).
2. OpenStack Cloud Software: Open source software for building private and public clouds. http://www.openstack.org. Febrero 2015.
3. OpenNebula. https://opennebula.org/. Febrero 2019
4. Galarza, B.; Tuamá, C.; Zaccardi, G.; Encinas, D.; Morales, M. “Implementaciones de Cloud Computing y aplicaciones en el ámbito universitario”. 1° Congreso Nacional de Ingeniería Informática y Sistemas de Información (CoNaIISI 2013). Ciudad de Córdoba, Argentina.
5. Zaccardi, G.; Galarza, B.; Encinas, D.; Morales, M. “Implementación de Cloud

Computing utilizando OpenStack". 2° Congreso Nacional de Ingeniería Informática y Sistemas de Información (CoNaIISI 2014). Ciudad de San Luis, Argentina.
6. Galarza, B.; Zaccardi, G.; Encinas, D.; Morales, M. "Análisis de despliegue de una IaaS utilizando Openstack". XXI Congreso Argentino de Ciencias de la Computación (CACIC 2015). Ciudad de Junín, Argentina.
7. OpenStack Deployment Fuel. https://www.mirantis.com/products/mirantis-openstack-software/openstack-deployment-fuel/. Febrero 2016
8. Amazon Web Services (AWS)-Cloud Computing Services. https://aws.amazon.com Marzo 2019
9. T. PVFS2, "PVFS 2 File System Semantics Document," tech. rep., PVFS Development Team, 2015
10. Atop Tool. https://www.atoptool.nl/index.php
11. Kreutz, D., Ramos, F. M. V., Esteves Verissimo, P., Esteve Rothenberg, C., Azodolmolky, S., & Uhlig, S. (2015). Software-Defined Networking: A Comprehensive Survey. Proceedings of the IEEE, 103(1), 14–76.
12. Mininet. http://mininet.org
13. Montes de Oca, F.; Galarza, B.; Morales, M.; Encinas, D. "Redes Definidas por Software en Entorno de Cloud Computing". 6° Congreso Nacional de Ingeniería Informática y Sistemas de Información (CoNaIISI 2018). Mar del Plata. Argentina.
14. Zaccardi, G.; Galarza, B.; Morales, M.;Encinas, D. "Despliegue y ejecución de un cloud privado". 4°Congreso Nacional de Ingeniería Informática y Sistemas de Información (CoNaIISI 2016). Ciudad de Salta, Argentina.
15. Armanno, M.; Navarro, G.; Morales, M.; Encinas, D. "Utilización de servicios de Cloud Computing y sensores". 8º Congreso Nacional de Ingeniería Informática/Sistemas de Información (CoNaIISI 2020). San Francisco, Córdoba.
16. González, N.; Lescano, N.; Pinto, C.; Morales, M.; Encinas, D. "Análisis de rendimiento de IaaS y PaaS". 8º Congreso Nacional de Ingeniería Informática/Sistemas de Información (CoNaIISI 2020). San Francisco, Córdoba.
17. Lustre Manual. https://doc.lustre.org/lustre_manual.xhtml#idm140436306123424. Febrero 2021.
18. Heichler, Jan. "An introduction to BeeGFS." (2014).
19. Benquerença Mendes, N.; Bond, R.; Morales, M.; Encinas, D. "Rendimiento de sistema de archivos en arquitecturas distribuidas y paralelas**".** 8º Congreso Nacional de Ingeniería Informática/Sistemas de Información (CoNaIISI 2020). San Francisco, Córdoba.
20. E. Párraga, B. León, R. Bond, D. Encinas, A. Bezerra, S. Mendez, D. Rexachs, E. Luque. (2021) Analyzing the I/O Patterns of Deep Learning Applications. Cloud Computing, Big Data & Emerging Topics. JCC-BD&ET 2021. Communications in Computer and Information Science, vol 1444. Springer, Cham. https://doi.org/10.1007/978-3-030-84825-5_1
21. Benquerença Mendes, N.; Bond, R.; Morales, M.; Encinas, D. "Sistema de Archivos Paralelos con Aplicaciones de Machine Learning.". XXVIII Congreso Argentino de Ciencias de la Computación (CACIC 2022). La Rioja, Argentina.
22. Rodrigo Lago; Verónica Scholz; Roman Bond; Martin Morales; Diego Encinas. Experiencia de desarrollo de una aplicación de reconocimiento de canciones mediante la técnica de huella de audio". En XXIX Congreso Argentino de Ciencias de la Computación CACiC, Argentina. 2023.
23. Alan Dell Orto Marengo, Emanuel Julio, Diego Rolón, Román Bond, Martin Morales, Diego Encinas. Simulación de sistema de reciclaje mediante banda clasificadora. En 11° Congreso Nacional de Ingeniería Informática y Sistemas de la Información, Argentina. 2023.
24. Emiliano Camer; Martin Páez; Nicolás Vitek; Román Bond; Daniel Martín Morales; Diego Encinas. "Estudio de simulación de procesos aplicado a la optimización de sistemas de producción". En XXX Congreso Argentino de Ciencias de la Computación CACiC, Argentina. 2024.

25. Albornoz, G.; Bond, R.; Morales, M.; Encinas, D. “Implementación de un Sistema de Reconocimiento Emocional Facial con Python y OpenCV”. 12º Congreso Nacional de Ingeniería Informática/Sistemas de Información (CoNaIISI 2024). Catamarca.
26. Yamila Ferrufino; Román Bond; Daniel Martín Morales; Diego Encinas. “Evaluación de la Eficiencia Energética y el Impacto de las Aplicaciones de Redes Sociales”. En XXX Congreso Argentino de Ciencias de la Computación CACiC, Argentina. 2024.

Arquitecturas Multiprocesador en HPC: Aspectos de Portabilidad, Eficiencia Energética y Computación Cuántica

Armando De Giusti [(1)(2)], Marcelo Naiouf[(1)], Fernando G. Tinetti [(1)(3)], Horacio Villagarcía [(1)], Franco Chichizola[(1)], Laura De Giusti[(1)(3)], Enzo Rucci[(1)(3)], Adrián Pousa[(1)], Victoria Sanz [(1)(3)], Diego Montezanti [(1)], Diego Encinas [(1)], Ismael Rodríguez[(1)], Sebastián Rodríguez Eguren[(1)], Leandro Libutti[(1)], Manuel Costanzo [(1)], César Estrebou[(1)], Javier Balladini[(4)]

[1]Instituto de Investigación en Informática LIDI (III-LIDI),
Facultad de Informática, Universidad Nacional de La Plata – Comisión de Investigaciones Científicas de la Provincia de Buenos Aires
[2]CONICET – Consejo Nacional de Investigaciones Científicas y Técnicas
[3]CICPBA – Comisión de Investigaciones Científicas de la Provincia de Buenos Aires
[4]Universidad Nacional del Comahue

{degiusti,mnaiouf,fernando,hvw,francoch,ldgiusti,erucci,apousa,vsanz,dmontezanti,dencinas,ismael,seguren, llibutti,mcostanzo,cesarest}@lidi.info.unlp.edu.ar; javier.balladini@gmail.com

Resumen

El eje de esta línea de I/D lo constituye el estudio de las arquitecturas multiprocesador que integran sistemas distribuidos y paralelos. Incluye como temas centrales:

- Arquitecturas many-core (GPU, procesadores MIC, TPUs), FPGAs, híbridas (diferentes combinaciones de multicores y aceleradores), asimétricas, y cuánticas.
- Desarrollo y evaluación de algoritmos paralelos sobre nuevas arquitecturas y su evaluación de rendimiento computacional y energético.
- Desarrollo y evaluación de algoritmos cuánticos. Comparación con algoritmos paralelos clásicos.
- Estudio y optimización de código heredado.
- Desarrollo y evaluación de estrategias de resiliencia.
- Modelado y simulación de E/S en HPC.

Palabras clave: *Sistemas Paralelos. Asimétricos. GPU, FPGA, TPU. Computación cuántica. Eficiencia energética. Resiliencia. Portabilidad. E/S paralela.*

Contexto

La línea de investigación que se presenta es parte del proyecto “Computación de Alto Desempeño y Distribuida: Arquitecturas, Algoritmos, Tecnologías y Aplicaciones en HPC, Fog-Edge-Cloud, Big Data, Robótica, y Tiempo Real.” del III-LIDI y de proyectos específicos apoyados por organismos nacionales e internacionales. También es parte del proyecto “Planificación eficiente de algoritmos de ML en contenedores utilizando multicores y GPUs” financiado por la Facultad de Informática de la UNLP.

En los temas relacionados existen cooperaciones con varias Universidades de Argentina, y se está trabajando con Universidades de América Latina y Europa en proyectos financiados por ERASMUS, CyTED y la OEI (Organización de Estados Iberoamericanos). En particular, se colabora con el proyecto “Computación de Altas Prestaciones Eficiente y Segura para Aplicaciones de Servicios de Salud Inteligentes”, de la Universidad Autónoma de Barcelona.

Por otra parte, se cuenta con financiamiento de diferentes empresas de Argentina, en particular para la formación de recursos humanos en las temáticas de Cloud, Fog y Edge Computing.

Se participa en iniciativas como el Consorcio de I+D+I en Cloud Computing, Big Data y Emergent Topics, conformado por Universidades de Argentina y España.

Asimismo, el III-LIDI forma parte del Sistema Nacional de Cómputo de Alto Desempeño (SNCAD) de la Secretaría de Innovación, Ciencia y Tecnología (SICYT)

Introducción

Una de las áreas de creciente interés lo constituye el cómputo de altas prestaciones, en el cual el rendimiento está relacionado con dos aspectos: por un lado, las arquitecturas de soporte, y por otro, los algoritmos que hacen uso de éstas.

A la aparición de arquitecturas *many-core* (como las GPU o los procesadores MIC), se ha sumado el uso de FPGAs, debido a su potencia de cómputo y rendimiento energético. Su combinación en sistemas HPC da lugar a plataformas híbridas con diferentes características [22]. Lógicamente, esto trae aparejada una revisión de los conceptos del diseño de algoritmos paralelos (incluyendo los mismos lenguajes de programación y el software de base), así como la evaluación de las soluciones que éstos implementan.

Si bien la heterogeneidad del hardware permite maximizar las ganancias de rendimiento y eficiencia energética, implica el uso de lenguajes, librerías y herramientas específicas para cada una de ellas, lo que aumenta el costo de desarrollo y mantenimiento del software [3].

GPUs y Cluster de GPUs

Las GPUs son el tipo de acelerador dominante en la comunidad de HPC hoy en día debido a su alto rendimiento y bajo costo de adquisición. En los últimos años, las GPUs han evolucionado hacia arquitecturas heterogéneas, que combinan diferentes clases de instrucciones. El principal desafío reside en cómo aprovechar esta vasta capacidad computacional de forma eficiente. En ese sentido, la tecnología Multi-Instance GPU (MIG) de NVIDIA representa un avance para los centros de datos [25].

Asimismo, la combinación de GPUs con otras plataformas paralelas, como clusters y multicores, brindan un vasto conjunto de posibilidades de investigación en arquitecturas híbridas. Los desafíos que se plantean aquí son múltiples, sobre todo en lo referido a distribución de datos y procesos en tales arquitecturas híbridas, a fin de optimizar el rendimiento de las soluciones.

FPGAs

Una FPGA (*Field Programmable Gate Array*) es una clase de acelerador basado en circuitos integrados reconfigurables. La capacidad de adaptar sus instrucciones de acuerdo con la aplicación objetivo le permite incrementar la productividad de un sistema y mejorar el rendimiento energético para ciertos tipos de aplicaciones. En los últimos años, se ha extendido su uso debido a la integración de estos dispositivos en arquitecturas híbridas [10][11]; al desarrollo de nuevas herramientas de programación, que ayudan a reducir los tradicionales tiempos y costos de programación [27][31]; y a la incorporación de FPGAs a los servicios de Cloud, lo que abre nuevas oportunidades para la explotación de esta clase de aceleradores.

TPUs

Las unidades de procesamiento tensorial (TPU) son una clase de Circuitos Integrados de Aplicación Específica (ASIC) desarrolladas por Google con el propósito de acelerar las cargas de trabajo de aprendizaje automático que requieren las aplicaciones desarrolladas en su framework TensorFlow [29]. Su uso provee una alternativa a otras arquitecturas ya conocidas como CPUs, GPUs y MICs. En ese sentido, interesa analizar las tasas de aceleración y eficiencia energética provistas por esta nueva arquitectura, en comparación con el resto.

Computación cuántica

La Computación Cuántica (QC) surge como un nuevo paradigma que aprovecha propiedades de la mecánica cuántica en computación. La QC posee características inherentemente paralelas y se espera que resuelva algunos problemas de forma más rápida que la computación clásica. Varios autores coinciden que la QC deberá integrarse al cómputo clásico HPC y esto abre nuevos desafíos [37][38].

Eficiencia energética

La mejora de la eficiencia energética es una de las principales preocupaciones en la informática actual, principalmente a partir de

las plataformas con gran cantidad de procesadores. Muchos esfuerzos están orientados a tratar la eficiencia energética y a las metodologías para medirla como ejes de I/D, como una métrica de evaluación relevante. Entre los puntos de interés se pueden mencionar la evaluación de eficiencia energética de diferentes algoritmos y plataformas paralelas, y su posterior optimización a través del estudio y aplicación de técnicas que reducen el consumo energético a nivel de aplicación, middleware o sistema.

Código heredado

La mayoría de los programas de simulación numérica que se emplean hoy en día fueron desarrolladas cuando las arquitecturas paralelas no existían. Por ello, este conjunto de aplicaciones abre oportunidades para desarrollar técnicas y herramientas que permitan optimizar el código, tanto desde el punto de vista computacional como desde la ingeniería de software [28].

Resiliencia

En la actualidad, lograr sistemas resilientes resulta un verdadero desafío considerando el creciente número de componentes, la cercanía a los límites físicos en las tecnologías de fabricación y la complejidad incremental del software. La corrección de las aplicaciones y la eficiencia en su ejecución se torna más importante en HPC debido a los extensos tiempos de ejecución. En ese sentido, resulta relevante desarrollar estrategias de detección y recuperación de fallos, especialmente a través de librerías de software [15]. Por otro lado, resulta de interés la aplicación de mecanismos de resiliencia en aplicaciones distribuidas [21].

Simulación y Entrada/Salida paralela

A pesar de los avances tecnológicos, las operaciones de E/S en los centros de supercómputo siguen siendo un cuello de botella para determinadas aplicaciones HPC. El rendimiento de un sistema depende de la carga de trabajo (patrones de E/S de las aplicaciones) y de su configuración (hardware y software) [19]. Contar con herramientas que permitan modelar y predecir el comportamiento de este tipo de aplicaciones en HPC resulta fundamental para mejorar su rendimiento [5].

El análisis y diseño de modelos de simulación basados en la arquitectura de E/S paralela, permite disminuir la complejidad y cubrir las exigencias de las aplicaciones en HPC, al poder identificar y evaluar los factores que influyen en las prestaciones [6]. Además. considerando la relevancia que ha cobrado la simulación de transmisión de enfermedades intra-hospitalarias, resulta de interés utilizar la experiencia desarrollada para el modelado de estos sistemas [13][32].

Dispositivos de bajo costo con capacidades para cómputo paralelo

En la actualidad se comercializan placas de bajo costo como Raspberry PI [20] u Odroid [33], que poseen múltiples núcleos simples. Asimismo, existen diversos dispositivos móviles con capacidades similares. Es de interés estudiar la manera de explotar el paralelismo en estos dispositivos para mejorar el rendimiento y/o consumo energético de las aplicaciones [1].

Líneas de Investigación, Desarrollo e Innovación

- Arquitecturas many-core (CPU, GPU y TPU) y FPGA. Análisis de este tipo de máquinas y de técnicas para desarrollar código optimizado.
- Arquitecturas híbridas (diferentes combinaciones de clusters, multicores, manycores y FPGAs). Diseño de algoritmos paralelos sobre las mismas. Técnicas de resiliencia.
- Exploración de nuevos lenguajes, modelos y estándares de programación para HPC.
- Consumo energético en las diferentes arquitecturas de alto desempeño, en particular en relación con los algoritmos paralelos y la configuración de la arquitectura. Modelado y estimación del consumo de potencia de arquitecturas HPC y distribuidas con capas de Cloud, Fog y Edge Computing.

- Análisis y desarrollo de modelos e implementación de simuladores de la pila de software de E/S en HPC.

Resultados y Objetivos

Investigación experimental por realizar

- Desarrollar y evaluar algoritmos paralelos sobre nuevas arquitecturas paralelas. Analizar rendimiento, eficiencia energética, costo de programación y portabilidad.
- Desarrollar y evaluar algoritmos en máquinas cuánticas. Comparar con algoritmos paralelos en máquinas tradicionales.
- Analizar las capacidades de lenguajes no convencionales para procesamiento paralelo, considerando rendimiento, costo de programación y portabilidad.
- Realizar el desarrollo de nuevos planificadores de tareas para multicores asimétricos sobre diferentes sistemas operativos con el objetivo de maximizar el rendimiento y minimizar el consumo de energía [24][25].
- Analizar proceso de migración de algoritmos entre arquitecturas diferentes, considerando portabilidad, rendimiento, esfuerzo de programación y portabilidad.
- Evaluar el consumo energético debido al procesamiento en los distintos niveles de la arquitectura distribuida y a las comunicaciones entre ellos
- Desarrollar técnicas de tolerancia a fallas que permitan aumentar la resiliencia de sistemas paralelos y distribuidos.
- Calibrar y sintonizar el modelado y simulación de E/S en HPC conseguido para distintos escenarios e infraestructuras.

Resultados obtenidos

- Usando una implementación optimizada para arquitecturas x86 como base [18], se diseñó y desarrolló una librería C para cómputo paralelo de caminos mínimos en arquitecturas multicore [23].
- Se realizaron y analizaron modificaciones al framework TensorFlow para permitir la maleabilidad de hilos en CPU [12].
- Se realizó un estudio de migración de códigos CUDA a SYCL usando el ecosistema oneAPI para aplicaciones del ámbito biológico y se evaluó el esfuerzo de programación requerido, la portabilidad funcional y de rendimiento en arquitecturas GPU, CPU y CPU-GPU[2][3].
- Se analizaron los desafíos en la integración de unidades de procesamiento cuántico a sistemas HPC tradicionales [37][38].
- Se diseñó la herramienta SEDAR y evaluó su desempeño para detección y recuperación de fallos transitorios [15].
- Se implementaron y evaluaron diferentes estrategias de resiliencia en una arquitectura basada en microservicios de gran escala [21][26].
- Se desarrolló un simulador que permite predecir cuál es la mejor estrategia para ahorrar energía ante una falla permanente en un sistema HPC [16].
- Se realizó un primer análisis comparativo del consumo energético cloud vs edge computing [40].
- Se analizó la resiliencia de un servicio de emergencias hospitalarias ante situaciones extremas mediante el uso de simulación [14][35][36].
- Se analizó el comportamiento del Dengue en la ciudad de La Plata mediante simulación buscando estimar su impacto de propagación [13].
- Se desarrollaron técnicas de modelado y simulación de E/S en HPC que permiten predecir cómo los cambios realizados en los diferentes componentes afectan a la funcionalidad y al rendimiento del sistema, para posteriormente ser capaces de optimizar [5][6].
- Se analizó el impacto de la E/S en aplicaciones de aprendizaje automático profundo [19].
- Se exploró el uso de diferentes microcontroladores para aplicaciones de aprendizaje automático [7] [8].
- Se desarrolló una plataforma distribuida, escalable, híbrida y redundante que permite aprovechar recursos de cómputo basados en arquitecturas ARM a través de un servicio de cloud [9].
- Se evaluó el impacto de las comunicaciones en un cluster heterogéneo de placas RPi [30]

Organización de Eventos

En el año 2024 se han organizado las XII Jornadas de Cloud Computing, Big Data & Emerging Topics (JCC-BD&ET 2024) en Argentina, con participación de especialistas académicos del país y el exterior, además de empresas con experiencia en Cloud Computing [4][17]. En junio de 2025 se tiene prevista la organización de las XIII JCC-BD&ET.

Formación de Recursos Humanos

Dentro de la temática de la línea de I/D el último año se concluyeron 2 tesis de doctorado [34][39]. Al mismo tiempo se encuentran en curso 2 tesis de Doctorado en Ciencias Informáticas y 2 tesis de Maestría.

Además, se participa en el dictado de las carreras de Doctorado en Ciencias Informáticas, y Magíster y Especialización en Cómputo de Altas Prestaciones de la Facultad de Informática de la UNLP (acreditadas por la CONEAU con categoría A, B y A, respectivamente), por lo que potencialmente pueden generarse nuevas Tesis de Doctorado y Maestría, además de Trabajos Finales de Especialización.

Existe cooperación con grupos de otras Universidades del país y del exterior, y hay tesistas de diferentes Universidades realizando su Tesis con el equipo del proyecto.

Respecto a las carreras de grado, se dictan por parte de integrantes de la línea de investigación tres materias directamente relacionadas con los temas de ésta: “Taller de Programación sobre GPUs”, “Cloud Computing y Cloud Robotics” y “Conceptos y Aplicaciones en Big Data”.

Referencias

[1] P. S. Rodríguez Eguren, F. Chichizola, and E. Rucci, “Análisis del uso de un cluster de Raspberry Pi para cómputo de alto rendimiento”. Actas del XXIV Congreso Argentino de Ciencias de la Computación (CACIC 2018), ISBN: 978-950-658-472-6, págs. 134-144, 2018.

[2] Manuel Costanzo, Enzo Rucci, Carlos García-Sánchez, Marcelo Naiouf, Manuel Prieto-Matías. “Assessing opportunities of SYCL for biological sequence alignment on GPU-based systems.” J Supercomput (2024). https://doi.org/10.1007/s11227-024-05907-2

[3] Manuel Costanzo, Enzo Rucci, Carlos García-Sánchez, Marcelo Naiouf, Manuel Prieto-Matías. “Comparing Performance and Portability Between CUDA and SYCL for Protein Database Search on NVIDIA, AMD, and Intel GPUs”. 2023 IEEE 35th International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD). ISBN 979-8-3503-0548-7. Págs. 141-148. Porto Alegre, Brasil. doi: 10.1109/SBAC-PAD59825.2023.00023. Octubre 2023.

[4] M. Naiouf, L. De Giusti, F. Chichizola, y L. Libutti, Eds., Cloud Computing, Big Data and Emerging Topics: 12th Conference, JCC-BD&ET 2024, La Plata, Argentina, June 25–27, 2024, Revised Selected Papers, vol. 2189. en Communications in Computer and Information Science, vol. 2189. Cham: Springer Nature Switzerland, 2025. doi: 10.1007/978-3-031-70807-7.

[5] D. Encinas, S. Mendez, M. Naiouf, A. De Giusti, D. Rexachs del Rosario, and E. Luque, “An Agent-Based Model for Analyzing the HPC Input/Output System” International journal on advances in systems and measurements vol. 13, num. 3 & 4, págs. 192-202, 2020.

[6] D. Encinas et al., «Agent Based Simulation as an Efficient Way for HPC I/O System Tuning», en Computational Science – ICCS 2024, L. Franco, C. de Mulatier, M. Paszynski, V. V. Krzhizhanovskaya, J. J. Dongarra, y P. M. A. Sloot, Eds., Cham: Springer Nature Switzerland, 2024, pp. 129-136. doi: 10.1007/978-3-031-63759-9_16.

[7] César A. Estrebou, Martín Fleming, Marcos D. Saavedra, Federico Adra, Armando E. De Giusti. “Lightweight Convolutional Neural Networks Framework for Really Small TinyML Devices.”. Third International Conference on Smart technologies, Systems and Applications (SmartTech-IC), ISBN 978-3-030-99170-8, págs. 3-16, Cham, doi: 10.1007/978-3-030-99170-8_1, marzo de 2022.

[8] César Armando Estrebou, Marcos David Saavedra, Federico Adra, Martín Fleming. “TinyML for Small Microcontrollers.” X Jornadas de Cloud Computing, Big Data & Emerging Topics (La Plata, 2022), ISBN 978-950-34-2126-0, págs. 42-46, julio de 2022.

[9] David Petrocelli, Armando De Giusti, Marcelo Naiouf. “Collaborative, Distributed, Scalable and Low-Cost Platform Based on Microservices, Containers, Mobile Devices and Cloud Services to Solve Compute-Intensive Tasks.” European Conference on Parallel Processing 2021, ISBN 978-3-031-06156-1, págs. 545-548, Cham, doi: 10.1007/978-3-031-06156-1_47, junio de 2022.

[10] IBM. “IBM and Xilinx Announce Strategic Collaboration to Accelerate Data Center Applications”. Disponible en https://www-03.ibm.com/press/us/en/pressrelease/48074.wss

[11] Intel. “Intel Acquisition of Altera”. Disponible en http://intelacquiresaltera.transactionannouncement.com

[12] Leandro Ariel Libutti. “Thesis overview: Scheduling Elastic Machine Learning Process Through Containers”. Journal of Computer Science and Technology (ISSN 1666-6038), vol. 23, núm. 2, págs. e17-e17, doi: 10.24215/16666038.23.e17, octubre de 2023

[13] D. O. Encinas, J. Matto, J. Stella, L. Maccallini, y F. Romero, «Dengue Dynamics: Modelling Spread and Environmental Interactions», Journal of Computer Science and Technology, vol. 24, n.o 2, Art. n.o 2, oct. 2024, doi: 10.24215/16666038.24.e14.

[14] Mercedes Planas, Eva Bruballa, Emilio Luque, Dolores Rexachs, Francisco Epelde, Armando De Giusti. “Evaluación del tiempo crítico para realizar acciones preventivas en el Servicio de Urgencias Hospitalarias”. XXIX Congreso Argentino de Ciencias de la Computación – CACIC 2024. Octubre de 2024. En prensa.

[15] D. Montezanti, E. Rucci, A. D. De Giusti, M. Naiouf, D. Rexachs, and E. Luque, “Soft errors detection and automatic recovery based on replication combined with different levels of checkpointing”. Future generation computer systems (ISSN 0167-739X), vol. 113, págs. 240-254, doi. https://doi.org/10.1016/j.future.2020.07.003, 2020.

[16] Marina Morán, Javier Balladini, Dolores Rexachs, Enzo Rucci. “Exploring Energy Saving Opportunities in Fault Tolerant HPC Systems” Journal of Parallel and Distributed Computing (ISSN 0743-7315), págs. 104797, doi: 10.1016/j.jpdc.2023.104797, noviembre de 2023

[17] A. E. De Giusti, M. R. Naiouf, F. Chichizola, L. C. De Giusti, y L. A. Libutti, Short papers of the 12th Conference on Cloud Computing, Big Data & Emerging Topics (JCC-BD&ET 2024). Facultad de Informática (UNLP), 2024. doi: 10.35537/10915/171300.

[18] S. Calderón, E. Rucci, y F. Chichizola, «Enhanced OpenMP Algorithm to Compute All-Pairs Shortest Path on X86 Architectures», en Computer Science – CACIC 2023, P. Pesado, W. Panessi, y J. M. Fernández, Eds., Cham: Springer Nature Switzerland, 2024, pp. 46-61. doi: 10.1007/978-3-031-62245-8_4.

[19] E. Parraga, B. Leon, S. Mendez, D. Rexachs, R. Suppi, y E. Luque, «An Empirical Method for Processing I/O Traces to Analyze the Performance of DL Applications», en Cloud Computing, Big Data and Emerging Topics, M. Naiouf, L. De Giusti, F. Chichizola, y L. Libutti, Eds., Cham: Springer Nature Switzerland, 2025, pp. 74-90. doi: 10.1007/978-3-031-70807-7_6.

[20] Raspberry PI. https://www.raspberrypi.org/

[21] S. L. Suárez, E. Rucci, V. Betran, y D. M. Montezanti, «Implementando estrategias de resiliencia en una arquitectura basada en microservicios», en Actas del XXIX Congreso Argentino de Ciencias de la Computación (CACIC 2023), oct. 2023, pp. 467-488. Accedido: 24 de abril de 2024. [En línea]. Disponible en: http://sedici.unlp.edu.ar/handle/10915/164996

[22] Rucci, Enzo: “Evaluación de rendimiento y eficiencia energética en sistemas heterogéneos para bioinformática”. Tesis de Doctorado en Ciencias Informáticas (Facultad de Informática – UNLP). 2016.

[23] Jerónimo Lambre, Enzo Rucci. “Librería C para Cómputo Paralelo de Caminos Mínimos en Grafos sobre Arquitecturas Multicore”. XXIX Congreso Argentino de Ciencias de la Computación – CACIC 2024. Octubre de 2024. En prensa.

[24] Juan Carlos Saez, Adrian Pousa, Daniel Chaver, Fernando Castro, Manuel Prieto Matias: “ACFS: A Completely Fair Scheduler for Asymmetric Single-ISA Multicore Systems”. In: ACM SAC 2015 (The 30TH ACM/SIGAPP Symposium on applied computing). 2015.

[25] NVIDIA, “GPU NVIDIA de Múltiples Instancias. ”https://www.nvidia.com/es-la/technologies/multi-instance-gpu/

[26] Sergio Leonel Suarez, Enzo Rucci, Victor Betran, Diego Montezanti. “Extendiendo las Estrategias de Resiliencia de una Arquitectura Basada en Microservicios” XXIX Congreso Argentino de Ciencias de la Computación – CACIC 2024. Octubre de 2024. En prensa.

[27] Sean Settle: “High-performance Dynamic Programming on FPGAs with OpenCL”. In: IEEE High Performance Extreme Computing Conference. 2013.

[28] F. G. Tinetti, M. J. Perez, A. Fraidenraich, and A. E. Altenberg, “Legacy code and parallel computing: updating and parallelizing a numerical model”. The journal of supercomputing (ISSN 1573-0484), doi. 10.1007/s11227-020-03172-7, 2020.

[29] Google Inc. “Cloud Tensor Processing Unit (TPU)” Disponible en https://cloud.google.com/tpu/docs/tpus?hl=es-419

[30] O. C. Valderrama Riveros and F. G. Tinetti, “MPI Communication Performance in a Heterogeneous Environment with Raspberry Pi”, Advances in Parallel & Distributed Processing, and Applications. Transactions on Computational Science and Computational Intelligence., ISBN: 978-3-030-69984-0, págs. 451-460, doi. 10.1007/978-3-030-69984-0_33, 2021.

[31] Xilinx Inc. “SDAccel Development Environment”. [Online]. Disponible en http://www.xilinx.com/products/design-tools/software-zone/sdaccel.html

[32] Maccallini, L., Encinas, D. O., & Romero, F. (2021). An Approach to the Modeling and Simulation of Intra-Hospital Diseases. Journal of Computer Science and Technology, 21(2), e14. https://doi.org/10.24215/16666038.21.e14

[33] Odroid http://www.hardkernel.com Accedido 21 de marzo de 2016.

[34] M. Costanzo «Thesis Overview: Viability Study of SYCL as a Unified Programming Model for Heterogeneous Systems Based on GPUs in Bioinformatics», Journal of Computer Science and Technology, vol. 24, n.o 2, Art. n.o 2, oct. 2024, doi: 10.24215/16666038.24.e18.

[35] Mariela Rodriguez, Francesc Boixader , Francisco Epelde, Eva Bruballa, Armando De Giusti, Alvaro Wong, Dolores Rexachs, Emilio Luque. “Análisis de desempeño de Servicio de Urgencia Hospitalario con llegada creciente de pacientes”. XXIX Congreso Argentino de Ciencias de la Computación – CACIC 2024. Octubre de 2024. En prensa.
[36] M. Ansari Dogaheh, M. Taboada, F. Epelde, E. Luque, D. Rexachs, y A. Wong, «An ABMS COVID-19 Propagation Model for Hospital Emergency Departments», en Cloud Computing, Big Data and Emerging Topics, M. Naiouf, L. De Giusti, F. Chichizola, y L. Libutti, Eds., Cham: Springer Nature Switzerland, 2025, pp. 103-117. doi: 10.1007/978-3-031-70807-7_8.
[37] A. Pousa y V. Sanz, «Desafíos en la integración de QPUs a sistemas HPC clásicos», Memoria investig. ing. (Facultad Ing., Univ. Montev.), n.º 27, pp. 238–247, dic. 2024.
[38] Pousa, A., Sanz, V. M., Naiouf, M., & De Giusti, A. (2024). Trends in High Performance Computing and Quantum Computing. Journal of Computer Science and Technology, 24(2), e11. https://doi.org/10.24215/16666038.24.e11
[39] O. C. Valderrama Riveros, «Arquitectura de Desarrollo y Evaluación de Vehículos no Tripulados», Tesis de Doctorado en Ciencias Informáticas, Universidad Nacional de La Plata, 2024. Accedido: 30 de septiembre de 2024. [En línea]. Disponible en: http://sedici.unlp.edu.ar/handle/10915/170779
[40] De Giusti, A., Montezanti, D., Medina, S. y Libutti, L. (2024). Edge Computing vs Cloud Computing: Optimizando el Consumo Energético. XXIX Congreso Argentino de Ciencias de la Computación – CACIC 2024. Octubre de 2024. En prensa.

Gestión distribuida de datos para aplicaciones de IoT

Claudio Omar Biale [1], Selva Nieves Ivaniszyn [1], Rubén Luis María Castaño [1], Ariel Lutenberg [2]

[1] Departamento de Informática, Facultad de Ciencias Exactas, Químicas y Naturales, Universidad Nacional de Misiones.

[2] CONICET-GICSAFe, Laboratorio de Sistemas Embebidos, Facultad de Ingeniería, Universidad de Buenos Aires

claudiobiale@fceqyn.unam.edu.ar, selvai@fceqyn.unam.edu.ar, rcastano@fceqyn.unam.edu.ar, alutenb@fi.uba.ar

RESUMEN

En un sistema tradicional de Internet de las Cosas (IoT), los datos recopilados por los nodos sensores y actuadores son enviados a la nube para su almacenamiento y análisis, mientras que los nodos reciben como respuesta comandos o instrucciones de control que producen cambios en sus actuadores. Este enfoque conduce a una alta latencia en la comunicación, un flujo de datos ascendente alto y mayores costos en los centros de datos en la nube. Adicionalmente, muchos sistemas de Internet de las Cosas experimentan problemas de conectividad que provocan la pérdida de datos si no existe un almacenamiento local coordinado. Este proyecto propone desarrollar un sistema distribuido que combina almacenamiento al borde de la red con servicios en la nube para mitigar estos inconvenientes. Este enfoque reduce latencias, optimiza el uso del ancho de banda y asegura la continuidad operativa en escenarios con conectividad limitada. Los resultados esperados incluyen un prototipo funcional que demuestre la viabilidad de esta arquitectura en aplicaciones reales.

Palabras clave: IoT, almacenamiento distribuido, borde-nube, conectividad, latencia.

CONTEXTO

Este trabajo de investigación se desarrolla en el marco del proyecto "16/Q2875-TI" de la Universidad Nacional de Misiones denominado “Gestión distribuida de datos para aplicaciones de IoT”.

1. INTRODUCCIÓN

La proliferación de dispositivos IoT plantea retos significativos en la gestión de datos debido a la alta dependencia de servicios en la nube. En regiones rurales con conectividad limitada, esta arquitectura no garantiza la operatividad continua, resultando en pérdidas de datos y latencias elevadas. La necesidad de sistemas de almacenamiento eficientes, con capacidades de consulta en tiempo real y procesamiento distribuido, ha impulsado el desarrollo de nuevas arquitecturas y enfoques en la gestión de datos en el borde de la red y la nube.

Diferentes estudios han explorado soluciones para manejar eficientemente los datos en entornos distribuidos. Feather es un sistema jerárquico y geo-distribuido que permite consultas eficientes con una garantía de frescura configurable, que balancea la latencia, el ancho de banda y la carga en los

nodos al borde [1]. FogStore introduce un modelo de almacenamiento clave-valor geo-distribuido, diseñado para optimizar la latencia y la tolerancia a fallos en entornos al borde, permite crear réplicas cercanas a los clientes para baja latencia y otras en ubicaciones remotas para tolerancia a fallas. Las consultas de clientes cercanos acceden a réplicas con consistencia fuerte, mientras que los clientes remotos reciben datos con consistencia eventual [2]. VergeDB plantea un enfoque basado en la compresión y análisis de datos en el borde que permite optimizar el uso de recursos computacionales disponibles, capacidad de almacenamiento y ancho de banda de red [3].

Además de estos enfoques, otros sistemas han explorado distintas optimizaciones en el almacenamiento y procesamiento de datos IoT. IoTDB propone un modelo de base de datos específicamente diseñado para manejar series temporales en entornos IoT, ofreciendo alta eficiencia en la ingesta y consulta de datos [4]. ModelarDB introduce una técnica de compresión mediante modelos matemáticos lo que permite representar series temporales con funciones que reducen drásticamente el tamaño de los datos sin perder precisión significativa. Implementa un mecanismo de transferencia continua de datos comprimidos desde el borde a la nube, esto asegura una administración eficiente del ancho de banda. Además, permite definir series temporales derivadas, que evitan el almacenamiento redundante de datos al recalcularlos dinámicamente [5]. TorqueDB es un motor distribuido para consultas sobre series temporales en dispositivos al borde y en la niebla, utiliza ElfStore para almacenamiento duradero en los nodos al borde e InfluxDB para consultas de los clientes en los nodos en la niebla [6].

Respawn es un sistema de almacenamiento distribuido multi-resolución para series temporales que particiona los datos entre la nube y el borde para optimizar consultas de baja latencia. Almacena datos en alta resolución en el borde y migra versiones agregadas y de menor resolución a la nube, para ello utiliza técnicas de predicción y caché adaptativa. Su arquitectura minimiza la transferencia innecesaria de datos y prioriza la migración de bloques estadísticamente relevantes, esto asegura un acceso eficiente sin depender de sincronización constante con la nube [7]. AnyLog, por su parte, es una base de datos descentralizada para gestión y consulta de datos IoT en arquitecturas al borde y en la nube. Utiliza una capa de coordinación descentralizada basada en un ledger compartido, que puede implementarse con blockchain u otros mecanismos. Los nodos se registran y sincronizan automáticamente, y gestionan datos localmente sin necesidad de redistribución constante. Este modelo elimina la dependencia de una base de datos centralizada y permite escalabilidad horizontal [8].

Otros enfoques han desarrollado estrategias específicas para el manejo de datos distribuidos. TS-Cabinet introduce una estrategia de almacenamiento jerárquico para bases de datos de series temporales en entornos de nube-borde-dispositivo final. Al utilizar modelos de temperatura de datos y predicción de carga de trabajo, este sistema optimiza la migración de datos entre diferentes niveles de almacenamiento para mejorar la eficiencia en las consultas y reducir la sobrecarga de almacenamiento [9]. LTSS, por otro lado, se centra en mejorar el rendimiento de la ingesta y la consulta de datos en entornos de alta carga al borde, ofreciendo una arquitectura optimizada para consultas en series temporales, superando en

rendimiento a soluciones como SQLite y PostgreSQL en múltiples escenarios de evaluación [10]. AIDA-DB propone un modelo de gestión de datos distribuidos para entornos al borde y en la nube que garantiza transacciones multi-sitio eficientes con baja latencia y alta escalabilidad. Implementa replicación total, ejecución optimista y, certificación y recuperación centralizada en la nube [11].

La investigación del trabajo se alinea con los desafíos actuales de sistemas de bases de datos distribuidas y almacenamiento eficiente de datos de sensores, explorando estrategias que optimicen la escalabilidad y la consulta de datos en entornos dinámicos y descentralizados. Se espera contribuir al desarrollo de soluciones más eficientes y adaptativas en la gestión de datos en IoT.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El trabajo de investigación aborda los siguientes ejes principales:

1. Almacenamiento distribuido: diseño de subsistemas de almacenamiento local en nodos al borde y almacenamiento en objetos en la nube.
2. Migración y almacenamiento de datos: estrategias para la compresión y transferencia de datos desde el borde a la nube de manera eficiente.
3. Consulta y escalabilidad: desarrollo de un servicio despachador escalable horizontalmente que coordine consultas globales que involucren a nodos al borde y a un sistema de almacenamiento en objetos en la nube. El sistema debe permitir realizar también consultas locales al borde de la red.

3. RESULTADOS OBTENIDOS/ESPERADOS

Este trabajo de investigación tiene como objetivo general el desarrollo de un sistema de captura, almacenamiento y consulta de datos distribuido transparente al usuario final donde los datos residen al borde de la red y en un servicio de almacenamiento en objetos en la nube. Las solicitudes al sistema pueden provenir de usuarios locales a un nodo al borde o de usuarios conectados a un servicio despachador ubicado en nodos en la nube.

En la actualidad, se encuentra en desarrollo el sistema de almacenamiento que consiste en una base de datos embebida al borde y un sistema de almacenamiento en objetos en la nube.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo incluye a docentes e investigadores de las diferentes universidades. Como parte de este proyecto se encuentra en desarrollo una tesis de Maestría en Internet de las Cosas bajo la dirección del Dr. Ing. Ariel Lutenberg.

Adicionalmente se pretenden desarrollar actividades de extensión en el marco del ciclo de charlas denominado “AgriTech: soluciones IoT y tecnológicas para la agricultura” dictadas por integrantes del equipo de trabajo.

5. BIBLIOGRAFÍA

[1] Mortazavi, S. H., Salehe, M., Gabel, M., & de Lara, E. (2020). Feather: Hierarchical querying for the edge. In *2020 IEEE/ACM Symposium on Edge Computing (SEC)* (pp. 271-284). IEEE.

[2] Gupta, H., & Ramachandran, U. (2018). Fogstore: A geo-distributed key-value

store guaranteeing low latency for strongly consistent access. In *Proceedings of the 12th ACM International Conference on Distributed and Event-based Systems* (pp. 148-159).

[3] Paparrizos, J., Liu, C., Barbarioli, B., Hwang, J., Edian, I., Elmore, A. J., Franklin, M. J., & Krishnan, S. (2021). VergeDB: A database for IoT analytics on edge devices. In *Conference on Innovative Data Systems Research (CIDR 2021)*.

[4] Wang, C., Qiao, J., Huang, X., Song, S., Hou, H., Jiang, T., Rui, L., Wang, J., & Sun, J. (2020). Apache IoTDB: Time-series database for internet of things. *Proceedings of the VLDB Endowment*, *13*(12), 2901-2904.

[5] Jensen, S. K., Thomsen, C., & Pedersen, T. B. (2023). ModelarDB: Integrated model-based management of time series from edge to cloud. In *Transactions on Large-Scale Data-and Knowledge-Centered Systems LIII* (pp. 1-33). Springer Berlin Heidelberg.

[6] Garg, D., Shirolkar, P., Shukla, A., & Simmhan, Y. (2020). TorqueDB: Distributed querying of time-series data from edge-local storage. In *Euro-Par 2020: Parallel Processing: 26th International Conference on Parallel and Distributed Computing, Warsaw, Poland, August 24–28, 2020, Proceedings 26* (pp. 281-295). Springer International Publishing.

[7] Buevich, M., Wright, A., Sargent, R., & Rowe, A. (2013). Respawn: A distributed multi-resolution time-series datastore. In *2013 IEEE 34th Real-Time Systems Symposium* (pp. 288-297). IEEE.

[8] Nawab, F., & Shadmon, M. (2024). The tipping point of edge-cloud data management. In *Conference on Innovative Data Systems Research (CIDR 2024)*.

[9] Cui, S., Wang, H., Liu, X., Tian, Z., & Ding, X. (2023). TS-Cabinet: Hierarchical storage for cloud-edge-end time-series database. *arXiv preprint arXiv:2302.12976*.

[10] Waddington, D. G., & Lin, C. (2016). A fast lightweight time-series store for IoT data. *arXiv preprint arXiv:1605.01435*.

[11] Faria, N., Costa, D., Pereira, J., Vilaça, R., Ferreira, L., & Coelho, F. (2022). AIDA-DB: A data management architecture for the edge and cloud continuum. In *2022 IEEE 19th Annual Consumer Communications & Networking Conference (CCNC)* (pp. 1-6). IEEE.

PREVENCIÓN DE ENFERMEDADES EN EL CULTIVO DEL ARROZ APLICANDO INTELIGENCIA ARTIFICIAL Y COMPUTACIÓN DE ALTO DESEMPEÑO

Miriam Asselborn[4], Julian Escalante[1], Mariela Lopresti [2], Natalia Miranda [2], Esteban Schab [1,3], Karina Cedaro[1], Malvina Martinez[5], Carlos Casanova[1,3], Virginia Pedraza[4], Fabiana Piccoli [1, 2]

[1] Universidad Autónoma de Entre Ríos, Facultad de Ciencia y Tecnología, Concepción del Uruguay
[2] LIDIC- Univ. Nacional de San Luis, San Luis
[3] Univ. Tecnológica Nacional, Facultad Regional Concepción del Uruguay, Concepción del Uruguay
[4] INTA, Estación Experimental Agropecuaria Concepción del Uruguay, Concepción del Uruguay
[5]Instituto de Clima y Agua. CIRN. INTA Castelar, Buenos Aires.

Argentina
{omlopres,ncmiran, mpiccoli}@unsl.edu.ar
{schabe, casanovac}@frcu.utn.edu.ar

RESUMEN

El arroz es uno de los cereales considerado un alimento básico a nivel mundial. En Argentina se cultiva en la región del litoral. Es un cereal propenso a enfermedades, la más común es el Quemado del Arroz (Pyricularia Oryzae). Predecir la aparición de la enfermedad permitiría tomar acciones preventivas y evitar pérdidas parciales o totales del cultivar. Este problema constituye un sistema complejo con gran volumen de datos. Una buena solución computacional demanda una importante cantidad de recursos a fin de lograr resultados en un plazo razonable.

En este trabajo se presenta una línea de investigación, la cual está centrada en el análisis de las condiciones predisponentes para detectar/predecir el riesgo de la aparición de la enfermedad y su propagación, permitiendo al agricultor tomar acciones preventivas. En esta línea se utiliza Inteligencia Computacional y Computación de Alto Desempeño para lograr la calidad y velocidad requeridas.

Palabras clave: Sistemas Complejos, Inteligencia Computacional. Computación de Alto Desempeño.

CONTEXTO

Esta propuesta de trabajo se lleva a cabo dentro de los proyectos de investigación: "Inteligencia Computacional aplicada al desarrollo de un modelo Eficiente de predicción de las condiciones predisponentes al Quemado del Arroz (Pyricularia Oryzae) en Argentina" (PICTO-UADER-UNER-00020, UADER e INTA) y "Tecnologías Avanzadas aplicadas al Procesamiento de Datos Masivos" (LIDIC, UNSL).

1. INTRODUCCIÓN

El uso de herramientas y técnicas computacionales permite llevar a cabo el estudio de grandes y/o complejos sistemas, permitiendo el análisis de posibles resultados a fin de evaluar la toma de decisiones y sus consecuencias. El modelado computacional es una herramienta para poder generar nuevos conocimientos en diferentes áreas de la ciencia tal como biología, física, agronomía, entre otras [11, 18, 24]. Incluir en la solución técnicas de inteligencia computacional, simulación y computación de alto desempeño [4, 25] es una buena opción.

Estas técnicas nos permiten desarrollar sistemas complejos, ya sea por las características de la solución o el tamaño/cantidad de los datos con los que se trabaja.

Generalmente, los sistemas complejos reales no son deterministas ni dicotómicos, con frecuencia requieren de gran cantidad de datos, imposibles de procesar y comprender al mismo tiempo por un humano [9, 27, 28]. Un modelo computacional de estos sistemas debe maximizar su utilidad, logrando un balance apropiado entre: complejidad, credibilidad e incertidumbre [19]. La Inteligencia Computacional (IC), también conocida como Soft Computing, se inspira en la naturaleza para desarrollar sistemas inteligentes basados en computadoras. Su estrategia es explotar la tolerancia a la incertidumbre, aproximación y falta de información para lograr eficiencia, robustez y soluciones de bajo costo. Las Redes Neuronales Artificiales, los Sistemas Difusos y la Computación Evolutiva son ejemplos de IC [7]. Estos métodos pueden ser utilizados sólo o en combinación para resolver un problema [23]. En un contexto de toma de decisiones, la IC tiende a soportar el proceso aprovechando la imprecisión e incertidumbre del mismo a fin de brindar "soluciones más satisfactorias". Las decisiones se derivan de analíticas, las cuales generan el conocimiento necesario según los datos disponibles. El diseño e implementación computacional de una analítica no es una tarea simple, se deben considerar, por ejemplo, aspectos como la toma de decisiones multicriterio y multipersona, la fusión de información de diferentes fuentes y naturalezas, el análisis de datos e información, etc.

Por lo expuesto, la implementación de una solución computacional para un sistema complejo, mediante un modelo que soporte la toma de decisiones, demanda gran cantidad de recursos para obtener resultados en un tiempo adecuado. Por la naturaleza de las técnicas de IC y las características de los recursos informáticos actuales, incluir estrategias de computación de alto desempeño es mandatoria [8, 14, 17, 20, 21].

La computación de alto desempeño (HPC) permite combinar recursos informáticos para solucionar problemas complejos, los cuales resultan grandes para ser resueltos en un tiempo adecuado en una sola computadora. HPC utiliza técnicas de computación paralela para procesar grandes conjuntos de datos, datos masivos y multidimensionales, y obtener resultados a mayor velocidad con tiempos de respuesta menores a los secuenciales. Esta computación puede resolverse usando recursos locales, por ejemplo clusters homo o heterogéneos, supercomputadoras en la nube o como una combinación de ambos [14]. Ejemplos de grandes problemas resueltos aplicando técnicas de HPC son: la secuenciación de ADN, extraer conocimiento de operaciones bursátiles, o tomar decisiones de conducción en tiempo real según los miles de datos obtenidos desde radares o GPS.

Un sistema real complejo es el relacionado con la producción de arroz. El arroz es el segundo cereal básico más importante a nivel mundial después del trigo. En Argentina tiene importancia regional, se cultiva bajo riego por inundación en las provincias de Entre Ríos, Corrientes, Santa Fe, Chaco y Formosa [22]. Entre las enfermedades que pueden afectar al arroz se encuentra el Quemado del Arroz (QA), causado por el hongo Pyricularia Oryzae. Esta enfermedad es la más importante a escala mundial [3, 10]. Las variedades de arroz utilizadas en Argentina son susceptibles al patógeno. Por ello es la enfermedad más temida, su aparición es esporádica, pero puede causar pérdidas totales a nivel de lote [16], los lotes llegan a tener 300 Ha aproximadamente. El control químico, mediante aplicaciones de fungicidas registrados durante la floración o principio de llenado de granos, ayuda a disminuir los daños. No obstante, el patógeno

posee una gran facilidad para desarrollar resistencias a los productos utilizados para su tratamiento [5].

Los indicios de la enfermedad pueden aparecer en cualquier etapa de desarrollo. Su intensidad varía con el ambiente y con la susceptibilidad del cultivar. El síntoma más temido es la necrosis del raquis de panoja, afectando directamente al rendimiento. Puede ser el primer signo de la enfermedad detectado por el productor/asesor en el campo y, en general, suele ser demasiado tarde como para una aplicación exitosa de fungicidas [2, 3].

La intensidad del monitoreo de los lotes y las condiciones naturales: suelo, clima, viento, horas sol, entre otras, constituyen las condiciones predisponentes al QA, siendo ampliamente estudiadas a nivel mundial. Los factores más importantes en el desarrollo de la enfermedad son la temperatura, la humedad relativa, la nubosidad, y las horas de mojado de hoja [6]. En condiciones naturales de campo, las prácticas de manejo del cultivo, como la elección del cultivar, el riego y la fertilización nitrogenada, pueden alterar la precisión de las predicciones de aparición de la enfermedad basadas en datos meteorológicos. La relación entre factores zonales y locales, y la intensidad de enfermedad, están siendo estudiadas para el desarrollo de modelos de predicción en Entre Ríos [16]. La multiplicidad de parámetros hace que el estudio de las interacciones de los mismos sea compleja.

Por todo lo expuesto, y según la bibliografía existente, el desarrollo del QA constituye un sistema candidato a ser estudiado mediante métodos computacionales de distinta índole. Se han reportado al menos 52 modelos predictivos, de los cuales sólo ocho se encuentran en uso actualmente en todo el mundo, y 4 de ellos han sufrido posteriores revisiones [6]. No obstante, nuevos modelos de predicción surgen debido a las dificultades que exhiben los existentes, sobre todo relativos a la calidad e integridad de los datos, y al alcance geográfico de los mismos [2, 12, 15, 26]. Sin embargo, existen aún muchas técnicas avanzadas de la IC que no se han probado en este problema y que pueden resultar superadoras respecto de las ya analizadas. Entre ellas se encuentran los autómatas celulares como modelos, la lógica difusa para la generación de reglas y otros tipos de arquitecturas de redes neuronales artificiales, así como las posibles hibridaciones entre tales técnicas. Todas ellas son adecuadas para la aplicación de HPC en su implementación. Es por ello que la línea de investigación se relaciona al desarrollo de un modelo de simulación para el problema del cultivo de arroz y los efectos del desarrollo del QA según las condiciones predisponentes. El modelo computacional se basa en autómatas celulares con toma de decisiones, aplicando técnicas de HPC en la solución.

2. LÍNEAS DE INVESTIGACIÓN, DESARROLLO E INNOVACIÓN

El productor agropecuario necesita tomar acciones inmediatas ante la real o posible presencia de QA en los cultivos, por ello es de utilidad contar con herramientas que le posibilite tomar decisiones y actuar en forma predictiva, adelantándose a los hechos o ante los primeros indicios de la enfermedad. Asimismo, se podría evitar aplicaciones preventivas innecesarias de fungicidas, con beneficios al medioambiente y a la economía de la producción.

Lograr una solución computacional para este tipo de sistema es complejo y multidisciplinar. El desarrollo de herramientas que permitirán analizar la interacción de distintos parámetros en un corto tiempo es una necesidad imperativa. Por todo esto, el desarrollo de un sistema computacional para la detección temprana del QA implica diferentes áreas de investigación, entre las cuales podemos destacar a:

- Procesamiento y visualización de datos meteorológicos.
- Modelos de Inteligencia Computacional.
- Computación de alto desempeño.

Cada una de estas líneas aportarán al desarrollo de una herramienta de software con emisión de alertas para un accionar preventivo, de manera inteligente, ante la presencia de condiciones predisponentes para el desarrollo de QA. Esto introduce no sólo valor agregado al sistema productivo arrocero, sino también genera conocimiento relacionado con el comportamiento de la enfermedad y las formas para su tratamiento, con la posibilidad de aplicar el modelo en entornos similares, ya sea en la misma disciplina u otra distinta.

3. RESULTADOS OBTENIDOS/ESPERADOS

Actualmente, se cuenta con un modelo de clasificación multinomial basado en regresión logística que utiliza datos meteorológicos para diagnosticar condiciones predisponentes a la aparición de la enfermedad en distintos niveles de impacto. No obstante, el modelo debe aún ser enriquecido utilizando datos extraídos *in situ* de lotes testigo que se refieren a la administración del cultivo: qué tipo de variedad de arroz es, si se aplicó o no fungicida, nivel de nitruración, entre otros.

Se está trabajando en la modelización del sistema mediante Autómatas Celulares con HPC y Agentes de Aprendizaje. Para ello se necesita el estudio previo del comportamiento de los cultivos y del QA, como así también de los parámetros que influyen en su desarrollo.

El sistema modela el ciclo del arroz usando autómata celular en GPU. Las condiciones predisponentes están definidas mediante parámetros, permitiendo su fácil variación para el análisis de las consecuencias. El sistema reproduce cada etapa del ciclo de crecimiento del cultivo y la posible aparición de la enfermedad en una de ellas.

Como uno de los objetivos es lograr soluciones paralelas portables, de costo predecible, capaz de explotar las ventajas de modernos ambientes HPC, los primeros desarrollos se realizan en GPU Nvidia con CUDA [13], y se prevé la incorporación de otras tecnologías como OpenCL [1] o HPC en la nube [14] .

4. FORMACIÓN DE RECURSOS HUMANOS

Los resultados esperados respecto a la formación de recursos humanos son el desarrollo de 2 tesis de doctorado y de varias tesinas de grado en las universidades intervinientes.

5. BIBLIOGRAFÍA

[1] M. Breyer, G. Daiss and D. Pflüger. “Performance - Portable Distributed k-Nearest Neighbors Using Locality - Sensitive Hashing and SYCL”. Int. Workshop on OpenCL. ISBN 9781450390330. Association for Computing Machinery. 2021

[2] Bregaglio, S., Titone, P., Cappellia, G, Tamborini, L., Mongiano, G., Confalonieri, R.. Coupling a generic disease model to the WARM rice simulator to assess leaf and panicle blast impacts in a temperate climate. Eur J Agron. 76:107–17. 2016.

[3] Carmona, M. Tizón del arroz. Herbario. UBA:http://herbariofitopatologia.agro.uba.ar/?page_id=5571. 2021.

[4] Huerta, E.A., Khan, A., Davis, E., Bushell, C., Gropp, W., Katz, D., Kindratenko, V., Koric, S., Kramer, W., McGinty, B., McHenry, K., Saxton, A.. Convergence of artificial intelligence and high performance computing on NSF-supported cyberinfrastructure. J Big Data 7, 88. Springer. 2020.

[5] Ishii, H., Hollomon, D.. Fungicide Resistance in Plant Pathogens : Principles and a Guide to Practical Management. Springer. ISBN 978-4431556411. 2015.

[6] Katsantonis, D., Kadoglidou. K., Dramalis, C., Puigdollers, P. Rice blast forecasting models and their practical value: a review. Phytopathologia Mediterranea, 56, 2. ISSN 0031-9465. Pp 187−216. 2017.

[7] Keller, J., Liu, D. Fogel, D.. Fundamentals of computational intelligence: Neural networks, fuzzy systems, and evolutionary computation. Wiley & Sons. ISBN 9781119214342. 2016.

[8] Kirk, D., Hwu, W.. Programming Massively Parallel Processors: A Hands-on Approach. Morgan Kaufmann. 3r ed. ISBN 978012811987. 2016

[9] Klir, G.. "Uncertainty and information: foundations of generalized information theory." Kybernetes . ISSN: 0368-492X, 35 (7-8). Pp. 1297-1299. 2006.

[10] Li, S., Zhang, X., Wang, W., Guo, X., Wu, Z., Du, W., ... & Xia, L.. Expanding the scope of CRISPR/Cpf1-mediated genome editing in rice. Molecular Plant, 11(7). Pp 995-998. 2018.

[11] Liu, J., Zheng, Q., Mu, X., Zuo, Y., Xu, B., Jin, Y., Li, Y.. Automated tumor proportion score analysis for PD-L1 (22C3) expression in lung squamous cell carcinoma. Scientific reports, 11(1). Pp 1-9. 2021.

[12] Nettleton, D. F., Katsantonis, D., Kalaitzidis, A., Sarafijanovic-Djukic, N., Puigdollers, P., Confalonieri, R. Predicting rice blast disease: machine learning versus process-based models. BMC bioinformatics. v 20, n 1, p. 1-16. 2019.

[13] Nvidia. “CUDA C++ Programming Guide”. https://docs.nvidia.com/cuda /pdf/ CUDA_C_Programming_Guide.pdf. 2023.

[14] Parra-González Ezra, Jaramillo Avila Uziel, Salazar-Linares P., Lara-Alvarez C. Tendencias y Desafíos de la Computación de Alto Rendimiento en la Nube. RISTI - Rev. Ibérica de Sistemas e Tecnologias de Informação. 131-146. 10.17013/risti.49.131-146. 2023.

[15] Pagani, V., Guarneri, T., Busetto, L., Ranghetti, L.,Boschetti, M., Movedi, E.,Campos-Taberner, M., Garcia-Haro, F., Katsantonis, D., Stavrakoudis, D., Ricciardelli, E., Romano, F.,Holecz, F., Collivignarelli, F., Granell, C., Casteleyn,S.,Confalonieri,R. A high-resolution, integrated system for rice yield forecasting at district level. Agricultural Systems, 168. Pp 181-190. ISSN 0308-521X. 2019.

[16] Pedraza, M.V.; Liberman, C.A.; Nuñez Bordoy, E.; Sanso, C.; Panozzo, J.; Asselborn, M.N. Control químico del Quemado del Arroz (Pyricularia oryzae). Jorn. Técn. Nac. del Cultivo de Arroz. Concordia, Entre Ríos. Pp 119-127. Resultados Experimentales 2012-2013. Vol XXII. Ed. Fundación Proarroz. Pp 67-74. ISSN 1853-8754. 2013.

[17] Piccoli, F.. *Computación de alto desempeño en GPU*.LaPlata: EDULP, ISBN:978-950-34-0759-2. 2011.

[18] Raja,R.; Nagwanshi,K.; Kumar, S.; Ramya Laxmi; K.. Data Mining and Machine Learning Applications. John Wiley & Sons. ISBN 9781119791782. 2022

[19] Ross, T.. Fuzzy logic with engineering applications. John Wiley & Sons, ISBN 9781119235866. 2016.

[20] Schab, E., Casanova, C., Piccoli, F.. Solving an Instance of a Routing Problem Through Reinforcement Learning and High Performance Computing.In: Cloud Computing, Big Data & Emerging Topics. JCC-BD&ET2022.Comm. in Computer and Information Science, vol 1634. Springer,Cham.https://doi.org/10.1007/978-3-031-14599-5_8. 2022.

[21] Schab, E., Lopresti, M. Miranda, N., Casanova, C., Piccoli, M. F.. “Toma de decisiones en sistemas de eventos mediante inteligencia computacional y computación de alto desempeño”. WICC 2022. RedUNCI. ISBN: 978-987-48222-3-9. Pp 535-539. Universidad Champagnat, Mendoza. Abril 2022.

[22] SIBER. Inf. Sup. Sembrada y Prod. del Arroz. Camp.2020/21.https://www.bolsacer.org.ar/Fuentes/siberd.php?Id=1259. 2021.

[23] Siddique, N.; Adeli, H.. Computational Intelligence: Synergies of Fuzzy Logic, Neural Networks and Evolutionary Computing. John Wiley and Sons. ISBN 9781118534816. 2013.

[24] Skoda, P.; Fathalrahman, A. Knowledge discovery in big data from astronomy and Earth observation : astro geoinformatics. Elsevier. ISBN 9780128191545. 2020.

[25] Wu, K., Simon, H.D.. High-Performance Computational Intelligence and Forecasting Technologies. Lawrence Berkeley National Laboratory. Ch: 22. Wiley. ISBN: 9781119482086. 2021.

[26] Yoshino R. Ecological studies on the penetration of rice blast fungus, Pyricularia oryzae into leaf epidermal cells. Bulletin of the Hokuriku National Agricultural Experiment Station. ;22: 163–221. 1979

[27] Zadeh, L. A.. "Outline of a New Approach to the Analysis of Complex Systems and Decision Processes," in IEEE Transactions on Systems, Man, and Cybernetics, vol. SMC-3, no. 1, pp. 28-44, 1973.

[28] Zimmermann, H. . Testability and meaning of mathematical models in social sciences. Mathematical Modelling 1(1). ISSN 0270-0255. Pp 123-139. 1980.

Procesamiento en Arquitecturas Distribuidas: Cloud, Fog y Edge Computing

Armando De Giusti [(1)(2)], Marcelo Naiouf[(1)], Santiago Medina, Diego Montezanti [(1)], Laura De Giusti[(1)(3)], Fernando G. Tinetti [(1)(3)], Franco Chichizola[(1)], Enzo Rucci[(1)(3)], Adrián Pousa[(1)], Victoria Sanz [(1)(3)], Diego Encinas [(1)], Ismael Rodríguez[(1)], Sebastián Rodríguez Eguren[(1)], Leandro Libutti[(1)], Manuel Costanzo [(1)], Francisco Garay[(1)], Federico Walas Mateo[(1)(4)]

[1]Instituto de Investigación en Informática LIDI (III-LIDI),
Facultad de Informática, Universidad Nacional de La Plata – Comisión de Investigaciones Científicas de la Provincia de Buenos Aires
[2]CONICET – Consejo Nacional de Investigaciones Científicas y Técnicas
[3]CICPBA – Comisión de Investigaciones Científicas de la Provincia de Buenos Aires
[4]Universidad Nacional Arturo Jauretche

{degiusti, mnaiouf, fernando, smedina, dmontezanti, ldgiusti, fernando, francoch, erucci, apousa, vsanz, dencinas, ismael, seguren, llibutti, mcostanzo, fgaray}@lidi.info.unlp.edu.ar, fedewalas@gmail.com

Resumen

Esta línea de I/D tiene como eje principal el estudio de arquitecturas distribuidas basadas en la integración de los paradigmas de Cloud, Fog y Edge Computing con la finalidad de almacenar, procesar y visualizar los datos generados por nodos sensores, para distintas aplicaciones que resultan de interés.

Los temas centrales de investigación son:

- La integración de servicios específicos para la gestión de datos.
- El despliegue y la utilización de plataformas de internet de las cosas.
- El análisis de estrategias para la distribución de la carga de procesamiento y almacenamiento de datos en los diferentes niveles de la arquitectura.
- El análisis de la integridad y la performance de las comunicaciones, según el grado de distribución del procesamiento.
- La comparación de los tiempos de respuesta utilizando cada uno de los niveles, en la búsqueda de optimizar la performance global de las aplicaciones.
- La búsqueda de la reducción del consumo energético, migrando un cierto nivel de inteligencia a los nodos finales.
- La incorporación de estrategias de resiliencia a la arquitectura distribuida y evaluación de su desempeño.
- El desarrollo de nodos sensores específicos para la recolección de datos en diferentes ambientes.
- El modelado de sistemas complejos de tiempo real y el desarrollo de software inteligente para la toma de decisiones, que pueda ejecutarse en distintos niveles de procesamiento.
- El modelado y desarrollo de gemelos digitales para aplicaciones de tiempo real (industria de procesos / gestión sanitaria / área agrícola) buscando tener respuestas predictivas inteligentes.

Palabras clave: *Arquitecturas Distribuidas. Cloud Computing. Fog Computing. Edge Computing. IoT. Aplicaciones de Tiempo Real. Eficiencia Energética. Resiliencia. Gemelos Digitales.*

Contexto

Esta línea de investigación forma parte del proyecto “Computación de Alto Desempeño y Distribuida: Arquitecturas, Algoritmos, Tecnologías y Aplicaciones en HPC, Fog-Edge-Cloud, Big Data, Robótica, y Tiempo Real” del III-LIDI y de proyectos específicos apoyados por organismos nacionales e internacionales.

Dentro de los temas relacionados a esta línea de investigación, existen cooperaciones con diferentes universidades de Argentina, además de otras universidades de América Latina y Europa, en proyectos financiados por ERASMUS, CyTED y la OEI (Organización de Estados Iberoamericanos). En particular, se realiza una colaboración con el proyecto

"Computación de Altas Prestaciones Eficiente y Segura para Aplicaciones de Servicios de Salud Inteligentes", de la Universidad Autónoma de Barcelona (UAB).

Se cuenta con financiamiento de diferentes empresas de Argentina, en lo referente a la formación de recursos humanos en las temáticas de *Cloud*, *Fog* y *Edge Computing*.

Por otro lado, el III-LIDI forma parte del Sistema Nacional de Cómputo de Alto Desempeño (SNCAD) del Ministerio de Educación, Cultura, Ciencia y Tecnología de la Nación.

Además, desde la línea de investigación y el proyecto principal se participa una iniciativa denominada "Consorcio de I+D+I en *Cloud Computing*, *Big Data* y *Emergent Topics*", conformado por diversas universidades de Argentina y España. Asimismo, se está trabajando en cooperación con la UAB (España), la UNNOBA y la UNAJ integrando proyectos específicos relacionados con las áreas de aplicación de Gemelos Digitales.

Introducción

La generación de grandes volúmenes de datos a partir de nodos sensores heterogéneos desplegados en diferentes entornos físicos, requiere de la implementación de arquitecturas distribuidas que sean capaces de soportar las cargas de procesamiento y almacenamiento, necesarios para cumplir con los tiempos de respuesta requeridos por las aplicaciones de tiempo real.

El ~~nuevo~~ modelo de arquitectura distribuida denominado "*Edge-Fog-Cloud Computing*" [1] busca la integración de diferentes niveles de procesamiento con diversos servicios, permitiendo a las aplicaciones beneficiarse en los siguientes aspectos:

- Reducir el volumen de comunicaciones hacia el Cloud, como también los tiempos de respuesta, mediante la realización de parte del cómputo en los mismos nodos sensores o en una capa intermedia (Edge o Fog).
- Permitir flexibilidad, mediante la posibilidad de trasladar mayor cantidad de procesamiento a los nodos sensores en función de la demanda y el contexto.
- Incorporar resiliencia a la respuesta del sistema, frente a fallos de comunicación o caídas de conexión, a partir de la redundancia de componentes en los distintos niveles.

Este modelo de arquitectura trae nuevos desafíos, como son [2-5]:

- La administración de los distintos niveles de almacenamiento de datos heterogéneos, manteniendo, a pesar de ello, la integridad y seguridad del sistema.
- La gestión de diferentes niveles de procesamiento que permitan la administración inteligente de recursos.
- La definición de protocolos y mecanismos de interoperabilidad entre las capas de la arquitectura y los servicios en la nube.
- El análisis e implementación de estrategias para tolerar diferentes tipos de fallos.
- La reducción de los tiempos de respuesta, a partir de la distribución optima de tareas en los distintos niveles.
- La implementación de estrategias para disminuir el consumo energético, incluyendo el análisis sobre la conveniencia de procesar en los distintos niveles [30].
- La evaluación comparativa de diferentes plataformas, aplicaciones y servicios en cada capa.
- El análisis de la escalabilidad con el aumento del número de nodos sensores y dispositivos finales.

Cloud Computing

El cómputo en la nube (*Cloud Computing*), simplifica el acceso y la utilización de grandes conjuntos de recursos físicos y lógicos (como infraestructura, plataformas de desarrollo, almacenamiento y/o aplicaciones), por medio de una interfaz de administración web, con un modelo de arquitectura "*virtualizada*" [6] [7]. Estos recursos se brindan como servicios ("*as a service*") y pueden reconfigurarse dinámicamente para adaptarse a una carga de trabajo variable (escalabilidad), logrando una utilización más flexible y elástica [8].

En una arquitectura distribuida, que integra desde nodos sensores hasta el cómputo en la nube, este último nivel se reserva para el procesamiento centralizado de algoritmos complejos que manejan grandes volúmenes de datos. Sin embargo, parte de estos datos pueden haber sido pre-procesados en las capas inferiores de *Edge* o *Fog* [9] [10].

Fog Computing

El paradigma de *Fog Computing* surge como respuesta al crecimiento de los desarrollos e implementaciones de IoT. Si bien estos desarrollos requieren procesamiento en la nube, algunas de sus características hacen que su implementación resulte compleja si únicamente se emplean tecnologías de *Cloud Computing* [10-12]. Por lo tanto, se implementa una capa de *Fog*, que está conformada por una plataforma intermedia que ofrece capacidades de cómputo, almacenamiento y servicios de comunicación entre los dispositivos *Edge* (conectados a los *End Devices* que adquieren los datos) y el *Cloud*.

Las características de *Fog Computing* [13-15] se orientan al manejo de:

- Aplicaciones distribuidas de tiempo real (es decir, de bajas latencias).
- Gran número de nodos heterogéneos, que puede escalar dinámicamente.
- Movilidad de los nodos (que utilizan comunicaciones predominantemente inalámbricas) y necesidad de conocer su ubicación.

Edge Computing

La necesidad de utilizar un modelo de *Edge Computing* surge del crecimiento exponencial de la cantidad de dispositivos sensores disponibles que cuentan con inteligencia local. Debido al continuo crecimiento de la IoT (y del tráfico que genera), resulta necesario realizar preprocesamiento cerca de los sensores, de modo de integrar datos locales y geográficamente distribuidos, para luego enviarlos a los niveles superiores (*Fog* o *Cloud*) [16-19] generando una menor sobrecarga debida a las comunicaciones. Las ventajas del modelo *Edge* son relativamente evidentes [10] [20] [21]:

- Posibilidad de dar respuesta al usuario en tiempo real, utilizando capacidades de procesamiento locales.
- Disminución del volumen de datos transmitidos.
- Mayor seguridad, obtenida a partir de conexiones privadas con servidores locales.
- Posibilidad de disminuir el consumo energético, debido a las características de los componentes y a la distribución del procesamiento.
- Mejora de la eficiencia global, mediante el uso de una arquitectura distribuida, débilmente acoplada.

Aplicaciones de tiempo real.

El modelo *Edge-Fog-Cloud Computing* es adecuado especialmente para su aplicación a problemas de tiempo real (*Cloud Robotics*, vehículos autónomos, emergencias sanitarias, monitoreos de salud personalizados, etc.), al permitir que las capas cercanas a los sensores y al usuario resuelvan parte del procesamiento en menor tiempo (y con menos comunicaciones), pudiendo responder de manera casi "inmediata", y reservando para el *Cloud* el procesamiento de los datos masivos *offline* (por ejemplo, un algoritmo sofisticado para entrenar un modelo de comportamiento que luego se traduce en mejoras en las tareas que se llevan a cabo en las capas *Edge* y *Fog* [22] [23]).

En particular, resulta de interés el modelado de sistemas complejos, tales como los de la industria de procesos, la agroindustria y la gestión de emergencias sanitarias, en los que resulta necesario recolectar datos de sensores en tiempo real y ejecutar software inteligente para la toma de decisiones.

Eficiencia energética

Actualmente, la mejora de la eficiencia energética es un tema central en la informática, principalmente a partir del surgimiento de plataformas con gran cantidad de procesadores orientadas al alto desempeño. Muchos esfuerzos están destinados a mejorar la eficiencia energética y desarrollar

metodologías de medición para obtener métricas de evaluación [23]. El modelo *Edge-Fog-Cloud* presenta mejoras en el consumo debido a dos factores:

- Los procesadores que se emplean en la capa Edge son habitualmente de bajo consumo y suelen contar con un modo sleep.
- La disminución de las comunicaciones produce la reducción del consumo energético asociado al tráfico de red.

Líneas de Investigación, Desarrollo e Innovación

- Estudio de diferentes arquitecturas distribuidas que permitan aplicar los modelos de Edge y Fog Computing.
- Administración e integración de recursos y datos en Edge y Fog Computing.
- Seguridad e integridad de los datos que se comunican entre niveles.
- Vinculación de los niveles Edge y Fog con el Cloud, utilizando plataformas y servicios específicos dedicados a IoT.
- Evaluación de estrategias de distribución de procesamiento y datos entre capas (por ejemplo, migrando cierta inteligencia al nivel Edge y a los nodos sensores), en la búsqueda de mejorar los tiempos de respuesta y reducir el consumo energético [30].
- Análisis de la escalabilidad de las aplicaciones sobre Edge-Fog-Cloud.
- Obtención de métricas de eficiencia considerando tiempos de respuesta / costo de comunicaciones / consumo energético.
- Definición de estrategias para incorporar resiliencia y control del consumo energético a las plataformas basadas en Edge, Fog y Cloud [18] [24].
- Diseño y despliegue de dispositivos de diferente conformación: nodos sensores, robots, drones, etc.
- Implementación de aplicaciones de tiempo real que aprovechan las tecnologías de Cloud, Fog y Edge Computing para controlar procesos de Transformación Digital en la industria y las organizaciones [25] [26].
- Análisis de la aplicación de Gemelos Digitales en la industria y los sistemas inteligentes de tiempo real.
- Investigación y exploración de herramientas de desarrollo de Gemelos Digitales para aplicaciones industriales.

Objetivos y Resultados Esperados

Investigación experimental por realizar

- Integración de plataformas IoT, servicios específicos y servicios provistos por Clouds públicos.
- Implementación de distintas arquitecturas distribuidas, donde se desplieguen diversos servicios y plataformas que se integren a las capas de Edge y Fog, optimizados para cada caso de estudio particular.
- Evaluación de diversos protocolos de comunicación, para la gestión de nodos de IoT.
- Análisis e implementación de estrategias de resiliencia [18] y control de consumo energético en aplicaciones que utilizan los niveles de Edge, Fog y Cloud.
- Definición de estrategias de balance de carga en aplicaciones que utilizan la arquitectura distribuida.
- Aplicación: red de nodos sensores inteligentes para monitoreo y control de consumo energético y diversos parámetros ambientales en diferentes entornos.
- Aplicación: exploración y mapeo de entornos mediante nodos finales constituidos por robots, monitoreados por dispositivos en la capa Edge.

Resultados obtenidos

- Se han desplegado y evaluado diversas plataformas IoT para la gestión de nodos sensores [27].
- Se han integrado diferentes servicios específicos orientados a la gestión de nodos IoT.
- Se ha realizado la integración de la plataforma Home Assistant con sensores de CO2, de temperatura, humedad y una cámara [31].
- Se ha desarrollado un nodo Edge (basado en Raspberry Pi) que implementa el control

de los movimientos de un robot incorporando resiliencia [32].

- Se han estudiado protocolos de comunicación, volúmenes de tráfico y tiempos de respuesta en aplicaciones distribuidas en tiempo real.
- Se ha desarrollado una plataforma a medida para la gestión de nodos que controlan el consumo energético y diversos parámetros ambientales en aulas.
- Se ha desarrollado un dispositivo final basado en un robot con cámara y diferentes sensores, conectado a un nodo Edge para su uso con diversas aplicaciones distribuidas.
- Se han implementado diversas estrategias de resiliencia sobre una arquitectura distribuida basada en microservicios [28].
- Se ha llevado a cabo una estadía Postdoctoral en el III-LIDI vinculado con Gemelos Digitales y su aplicación en la industria de procesos [29].

Organización de Eventos

En 2024 se han organizado las XII Jornadas de *Cloud Computing*, *Big Data* & *Emerging Topics* (JCC-BD&ET 2024) en Argentina, con participación de especialistas académicos del país y el exterior, además de empresas con experiencia en *Cloud Computing*.

Formación de Recursos Humanos

Dentro de la temática de la línea de I/D están en curso 1 Tesis de Doctorado y 1 Trabajo Final Integrador de Especialización.

Además, se están llevando a cabo 1 trabajo final de grado y 3 prácticas profesionales supervisadas.

Asimismo, se finalizaron 2 trabajos de grado y 3 prácticas profesionales supervisadas.

Se llevó a cabo en la Facultad de Informática de la UNLP una estadía Postdoctoral con el titulo “Automatización en la Industria X.0. Aplicación de Gemelos Digitales en industria de Procesos”, como una continuidad del trabajo realizado en [29]. Como resultado más saliente, se presentó el proyecto CyTED "Gemelos Digitales en la gestión predictiva e inteligente de servicios, productos, procesos o sistemas" junto a otras instituciones académicas de Iberoamérica. (Argentina, Uruguay, Chile, Colombia, Brasil, Ecuador y España).

Varios integrantes de la línea de investigación participan en el dictado de las carreras de Doctorado en Ciencias Informáticas, Magíster y Especialización en Cómputo de Altas Prestaciones y el Magíster en Gestión y Tecnología de Ciudades Inteligentes de la Facultad de Informática de la UNLP (acreditadas por CONEAU), por lo que potencialmente pueden generarse Tesis de Doctorado y Maestría, además de Trabajos Finales de Especialización.

Se dictaron tres cursos de postgrado: “Conceptos de *Cloud*, *Fog* y *Edge Computing*”, “Gobernanza y Gestión de Tecnologías de la Información” y “Tópicos de Fallos y Resiliencia en Arquitecturas Distribuidas”.

Además, algunos integrantes de esta línea de investigación dirigieron y participaron del proyecto “Procesamiento Eficiente de Grandes Datos mediante Cómputo de Altas Prestaciones, *Fog* y *Edge*”, financiado por la Facultad de Informática de la UNLP.

Se avanzó en la definición de una carrera de Maestría en Transformación Digital, en la que se requiere del dictado de cursos relacionados con el empleo de *Cloud*, *Fog* y *Edge Computing*.

Existe cooperación con grupos de otras Universidades del país y del exterior con posibilidad de realizar Tesis en colaboración.

Además, algunos integrantes de la línea de investigación participan en el dictado de dos asignaturas directamente relacionadas con los temas de ésta: “*Cloud Computing* y *Cloud Robotics*” y “Conceptos y Aplicaciones en *Big Data*”, dentro de las carreras de grado.

Asimismo, todos los años se llevan adelante proyectos de innovación y desarrollo con alumnos, relacionados con aplicaciones de tiempo real con redes de sensores, robots y drones controlados por nodos que operan en las capas de *Edge* y *Fog*.

Referencias

[1] Mohan N., Kangasharju J. “Edge-Fog Cloud: A Distributed Cloud for Internet of Things

Computations", 2016 Cloudification of the Internet of Things (CIoT), Paris, France, 2016, pp. 1-6.
[2] P. Garcia Lopez et al. "Edge-centric computing: Vision and challenges," *SIGCOMM Comput. Commun. Rev.*, vol. 45, no. 5, pp. 37–42, Sep. 2015.
[3] M. Yannuzzi et al. "Key ingredients in an iot recipe: Fog computing, cloud computing, and more fog computing," in *IEEE CAMAD*, 2014.
[4] K. Hong et al. "Mobile fog: A programming model for large-scale applications on the internet of things," in *ACM SIGCOMM Workshop on Mobile Cloud Computing*,2013.
[5] A. Chandra, J. Weissman, and B. Heintz, "Decentralized edge clouds," *Internet Computing, IEEE*, vol. 17, no. 5, pp. 70–73, Sept 2013.
[6] E. Rucci, M. Naiouf, F. Chichizola, and L. De Giusti "Cloud Computing, Big Data & Emerging Topics. 8th Conference, JCC-BD&ET 2020, La Plata, Argentina, September 8-10, 2020, Proceedings". Springer CCIS, ISBN: 978-3-030-61218-4, 2020.
[7] Xing, Y., Zhan, Y.: "Virtualization and Cloud Computing". In: Proceedings pp.305-312, Springer Link. ISBN 978-3-642-27323-0. (2012). Morgan Kaufmann. 2013.
[8] Velte, A.T., Velte, T.J., Elsenpeter, R.: "Cloud Computing: A Practical Approach". McGraw Hill Professional. 2009.
[9] Ashkan Yousefpour, Caleb Fung, Tam Nguyen, Krishna Kadiyala, Fatemeh Jalali, Amirreza Niakanlahiji, Jian Kong, Jason P. Jue. "All one needs to know about fog computing and related edge computing paradigms: A complete survey", Journal of Systems Architecture, Volume 98, 2019, Pages 289-330, ISSN 1383-7621,
[10] P. Lea, "IoT and Edge Computing for Architects: Implementing edge and IoT systems from sensors to clouds with communication systems, analytics, and security". Packt Publishing, 2020.
[11] Keyan Cao, Yefan Liu, Gongjie Meng, Quimeng Sun "An Overview on Edge Computing Research" in IEEE Access, vol. 8, pp. 85714-85728, 2020.
[12] W. S. Shi, X. Z. Zhang, and Y. F.Wang, "Edge computing: State-of-the-art and future directions," J. Comput. Res. Develop., vol. 56, no. 1, pp. 1_21, 2019.
[13] X. Hong and Y. Wang, "Edge computing technology: Development and countermeasures," *Chin. J. Eng. Sci.*, vol. 20, no. 2, p. 20, 2018.
[14] D. Evans. "The Internet of Things How the Next Evolution of the Internet is Changing Everything". Available:
https://www.researchgate.net/publication/30612290
[15] M. Satyanarayanan, "The emergence of edge computing," *Computer*, vol. 50, no. 1, pp. 30_39, Jan. 2017.
[16] Y. Q. Gao, H. Bguan, and Z. W. Qi, "Service level agreement-based energy-Efficient resource man agreement in cloud data centers," *Comput. Elect. Eng.*, vol. 40, no. 5, pp. 1621_1633, 2014.
[17] W. Shi, H. Sun, J. Cao, Q. Zhang, and W. Liu, "Edge computing-an emerging computing model for the Internet of everything era," *J. Comput. Res. Develop.*, vol. 54, no. 5, pp. 907_924, May 2017.
[18] Medina, S., Montezanti, D., Gómez D'Orazio, L., Garay, F., De Giusti, A., & Naiouf, M. (2023, June). Distributed Architectures Based on Edge Computing, Fog Computing and End Devices: A Conceptual Review Incorporating Resilience Aspects. In Conference on Cloud Computing, Big Data & Emerging Topics (pp. 31-44). Cham: Springer Nature Switzerland.
[19] F. Bonomi, R. Milito, J. Zhu and S. Addepalli, "Fog Computing and Its Role in the Internet of Things", in *MCC '12: Proceedings of the first edition of the MCC workshop on Mobile cloud computing.*
Association for Computing Machinery, New York,NY, United States, 2012.
[20] M. Asemani, F. Jabbari, F. Abdollahei and P. Bellavista, "A Comprehensive Fog-enabled Architecture for IoT Platforms", *High-Performance Computing and Big Data Analysis. TopHPC 2019. Communications in Computer and Information Science, vol 891. Springer, Cham*, 2019.
[21] W. S. Aung and S. Aung Nyein Oo, "Monitoring and Controlling Device for Smart Greenhouse by using Thinger.io IoT Server". *International Journal of Trend in Scientific Research and Development*, 2019.
[22] Dr. S. K. Selvaperumal, W. Al-Gumaei, R. Abdulla and V. Thiruchelvam, "Integrated Wireless Monitoring System Using LoRa and Node-Red for University Building". *Journal of Computational and Theoretical Nanoscience, Volume 16, Number 8, American Scientific Publishers*, 2019.
[23] M. Pi Puig, J. M. Paniego, S. Medina, S. Rodriguez Eguren, L. Libutti, J. Lanciotti, J. de Antueno, C. Estrebou, F. Chichizola and L. De Giusti, "Intelligent Distributed System for Energy Efficient Control", *Cloud Computing and Big Data. JCC&BD 2019. Communications in Computer and Information Science, vol 1050. Springer, Cham*, 2019.
[24] Medina, S., Montezanti, D. M., Gomez D'Orazio, L., Compagnucci, E., De Giusti, A. E., & Naiouf, M. (2022). Incorporating Resilience to Platforms based on Edge and Fog Computing. In X Jornadas de Cloud Computing, Big Data & Emerging Topics (La Plata, 2022).
[25] F. Walas, A.Redchuk, J.Tornillo. "Gestión de datos de la operación en la industria a través de plataformas IIoT. Experiencia en una empresa del sector alimentos en Argentina"
[26] Borek, A., Prill, N. "Driving Digital Transformation through Data and AI: A Practical Guide to Delivering Data Science and Machine Learning Products." Kogan Press, 2023.
[27] Gomez D´Orazio, L., Medina, S., Montezanti, D. "Integración de una red de sensores con una plataforma IoT para control inteligente de aulas". Proceedings del XXVIII Congreso Argentino de Ciencias de la Computación (CACIC 2022) – EUDELAR, pp. 1002 – 1007. ISBN 978-987-1364-31-2 (2022). http://sedici.unlp.edu.ar/handle/10915/149102.

[28] Suarez, Sergio Leonel. “Análisis de patrones de resiliencia en una arquitectura basada en microservicios”. Tesina de Licenciatura en Sistemas. Universidad Nacional de La Plata. 2022. http://sedici.unlp.edu.ar/handle/10915/149187
[29] Federico Walas (2023) Tesis Doctoral. “Nuevos modelos de negocio en el paradigma Industria 5.0. Inteligencia Artificial y Aprendizaje Automático para optimizar procesos industriales”
[30] De Giusti, A., Montezanti, D., Medina, S. y Libutti, L. (2024). Edge Computing vs Cloud Computing: Optimizando el Consumo Energético. In XXX Congreso Argentino de Ciencias de la Computación (CACIC)(La Plata, 7 al 11 de octubre de 2024).
[31] Dell´Arciprete, D., Medina, S., & Montezanti, D. M. (2024). Despliegue y análisis de la plataforma Home Assistant para la integración de nodos sensores heterogéneos. In XXX Congreso Argentino de Ciencias de la Computación (CACIC)(La Plata, 7 al 11 de octubre de 2024).
[32] Pérez Balcedo, O., Medina, S., & Montezanti, D. M. (2024). Estrategia para el control resiliente de vehículos robot utilizando Edge Computing para mapeo de entornos. In XXX Congreso Argentino de Ciencias de la Computación (CACIC)(La Plata, 7 al 11 de octubre de 2024).

Scheduling en Computación Serverless

Nelson Rodríguez[1], María Murazzo[1], Marcelo Moreno[1], Leonardo Celador [2], Ciro Delgado[2]

[1]Departamento e Instituto de Informática - F.C.E.F. y N. - U.N.S.J.

[2] Alumno Avanzado Licenciatura en Sistemas de Información y Cs. de la Computación - F.C.E.F. y N. - U.N.S.J.

Complejo Islas Malvinas. Cereceto y Meglioli. 5400. Rivadavia. San Juan, 0264 4234129

nelson@iinfo.unsj.edu.ar, marite@unsj-cuim.edu.ar, mpmoren@gmail.com, leonardomiguelcelador@gmail.com, cirodelgado7@gmail.com

Resumen

Serverless es una arquitectura que ha revolucionado la computación en el Cloud, pasando de la infraestructura tradicional basada en servidores a una de servicios totalmente gestionados y basados en eventos. Bajo el modelo Serverless, el proveedor de servicios asume la responsabilidad de las tareas de administración de la infraestructura, mientras que los desarrolladores pueden concentrarse en crear aplicaciones, escribiendo el mejor código de aplicación front-end y lógica empresarial. Todas las aplicaciones requieren servidores para ejecutarse, pero en el modelo serverless, un proveedor de servicios se encarga del trabajo rutinario; aprovisiona, escala y mantiene la infraestructura subyacente, realizando varias tareas, como la administración del sistema operativo, la aplicación de parches de seguridad, la administración del sistema de archivos y la capacidad, el balanceo de carga, la supervisión y el registro. De esta forma se obtienen muchos beneficios, como utilizar una infraestructura de servidores rentable, eficiente que permite una escalabilidad extrema, eliminando el costo del tiempo de inactividad. Un aspecto sumamente importante es la provisión de recursos o scheduling para la ejecución de funciones, las mismas se desarrollan considerando métricas similares al rendimiento, prioridad, latencia, costo, etc. Por lo tanto, el proceso de asignación de recursos se considera un factor crítico para los proveedores de servicios, que además impacta en la eficiencia y costos que pagan los usuarios. Dicha asignación debe realizarse a tiempo, elegir la técnica más apropiada para el tipo de aplicación y considerar las características propias de Serverless.

Palabras clave: *Serverless Computing, Scheduling, Serverless Function Scheduling, Cloud Computing.*

Contexto

El presente trabajo se encuadra dentro del área de I/D Procesamiento Distribuido y Paralelo y es la línea marco del proyecto de investigación Soporte Serverless para aplicaciones móviles de nueva generación, cuya propuesta se encuentra en ejecución para el período 2023-2024 y se extendió un año más Asimismo el grupo de investigadores viene trabajando en proyectos relacionados con la computación móvil, distribuida y de alta performance, desde hace más de 24 años, al que se han incorporado estudiantes de informática. A su vez se mantienen vínculos con investigadores de otras universidades, con los cuales se desarrollaron proyectos conjuntos en años anteriores, esto último favorece notablemente a todas las instituciones participantes.

Introducción

Serverless computing ha ganado impulso en los últimos años, dado que el 45% de las empresas ya implementan estas arquitecturas en entornos de producción y, además, otro 35% de la industria, experimenta activamente con estas tecnologías. Esto indica principalmente un profundo cambio de paradigma en la forma en que las organizaciones desarrollan aplicaciones y administran la infraestructura, dejando los modelos centrados en el servidor en favor de soluciones más flexibles y escalables [1].

En este paradigma emergente, las aplicaciones de software se descomponen en múltiples funciones independientes sin estado [2] [3]. Las funciones solo se ejecutan en respuesta a acciones desencadenantes (como interacciones de usuario, eventos de mensajería o cambios en la base de datos), y se pueden escalar de forma independiente y pueden ser efímeras (pueden durar una invocación) y están completamente administrados por el proveedor de Cloud.

Por otro lado, los algoritmos de scheduling se utilizan en serverless computing para asignar recursos, des esta forma, se pueden mejorar la utilización de dichos recursos, logrando optimizar parámetros como costos, eficiencia, uso de memoria, entre otros Este proceso puede ser difícil de lograr, debido a una serie de factores, como el comportamiento dinámico, los recursos heterogéneos, las cargas de trabajo que varían en volumen y las variaciones en el número de solicitudes [4].

El uso de funciones Serverless, a pesar de las ventajas, introducen un nuevo conjunto de desafíos en términos de scheduling y gestión de recursos para los proveedores [5]

El área de investigación de scheduling aplicado a Serverless ha tomado interés en estos últimos años Las publicaciones de referencia son recientes.

Por ejemplo la última revista de IEEE Internet Computing, trata sobre los nuevos desafíos de investigación y desarrollo en Serverless, que abarca desde el desarrollo de un simulador de asignación y programación para evaluar diversas políticas de orquestación, hasta el soporte para la implementación escalable y rentable de modelos de aprendizaje automático serverless en entornos periféricos heterogéneos [6].

Serverless

El término Cloud Computing entró en uso popular en 2008, aunque la práctica de proporcionar acceso remoto a las funciones informáticas a través de redes se remonta a los sistemas de tiempo compartido de los años 1960 y 1970 [7].

Amazon introdujo sus servicios minoristas basados en la web. Fue la primera gran empresa en pensar en usar solo el 10% de su capacidad (que era común en ese momento) como un problema a resolver [8] [9].

Posteriormente, y siguiendo la evolución de la contenerización, los servicios en la nube se han adaptado para ofrecer contenedores de mejor ajuste que requieren menos tiempo para cargar (arranque) y proporcionar mayor automatización en el manejo (orquestación) de contenedores en nombre del cliente, [9], surgiendo de esta forma la Computación Serverless.

Esta arquitectura impactó de tal manera, que se prevé que el tamaño del mercado de serverless computing crezca significativamente, alcanzando los 41 mil millones de dólares en 2028, en comparación con los 19 mil millones de dólares invertidos en 2024 [10]

El modelo de computación serverless es impulsado por eventos, los recursos se proporcionan como servicios escalables. En el modelo tradicional se cobra un costo fijo y recurrente por los recursos utilizados, independientemente de la cantidad de trabajo. En Serverless los clientes solo pagan por el uso del servicio y no se cobran costos asociados con el tiempo de inactividad.

Un modelo basado en funciones como Serverless, es adecuado para ráfagas, uso de CPU intensivo, cargas de trabajo granulares. Actualmente, los casos de uso de FaaS varían ampliamente, incluido el procesamiento de datos, el procesamiento de flujo, procesamiento de datos en tiempo real, la computación de borde (IoT) y la computación científica [11].

Una plataforma de función como servicio (FaaS) admite la creación de aplicaciones distribuidas compuestas por funciones Cloud pequeñas y de una sola tarea. Estas funciones se ejecutan en entornos de espacio aislado ligeros, que se ejecutan sobre máquinas virtuales. El proveedor de la nube administra los espacios aislados, los entornos de tiempo de ejecución y las máquinas virtuales.

Por lo tanto, un desarrollador puede crear aplicaciones elásticas en la nube sin tener que preocuparse por el aprovisionamiento del servidor y los administradores de elasticidad.

Existen diversos desafíos, oportunidades y problemas a resolver, entre ellos la experiencia del desarrollador [11], Interoperabilidad, testing, composición de funciones, vulnerabilidades de seguridad, ineficiencia en la gestión de recursos, administración del ciclo de vida, administración de requerimientos no funcionales, performance, optimización del overhead, mantenimiento del estado, cold start, composición de funciones, ingeniería para costo-performance, scheduling de recursos, ubicación para la ejecución de funciones, entre otros [12] [13] [14].

Scheduling en Serverless Computing

Scheduling en el contexto de Computación y Sistemas Operativos, se define como el proceso de asignación y administración de la ejecución de procesos o tareas de recursos de sistema, tales como CPU. El objetivo principal del scheduling es asegurar que los recursos son utilizados eficientemente, mientras mantiene el balanceo, capacidad de respuesta y throughput [15].

Otra definición indica lo siguiente: los problemas de scheduling exigen la tarea de asignar los recursos a los trabajos para lograr la función objetivo dentro de las restricciones especificadas de los recursos. En los problemas de scheduling clásicos, es necesario definir la secuencia para asignar trabajos en los recursos a medida que los recursos son fijos. Sin embargo, en el dominio de Cloud Computing, los proveedores de servicios pueden agregar y liberar recursos según los requisitos de la aplicación del usuario y, por lo tanto, controlar la cantidad de recursos que se utilizan para completar la tarea de ejecución de la aplicación [16].

El estudio del problema del scheduling de funciones serverless, ha surgido como una nueva área de investigación en los últimos tiempos. Se presentan varias soluciones en la literatura existente para el problema de encontrar un nodo host adecuado para programar una instancia de función, que pueda acomodar una sola solicitud o varias solicitudes simultáneas, en función de la arquitectura del sistema [17].

La aplicación de los algoritmos de scheduling en la arquitectura serverless, puede obedecer a resolver problemáticas específicas como: optimizar la computación serverless en el Edge para uso intensivo de datos [18], aplicaciones en ambientes multiproveedor [19], cargas de

trabajo en la continuidad Edge-Cloud [20].o reducir la latencia de la composición de funciones [21], también descarga y el scheduling inteligente entre otras [22].

Teniendo en cuenta los algoritmos de scheduling aplicables a la arquitectura de Serverless Computing, se pueden clasificar de acuerdo a su objetivo:

- scheduling del tiempo de ejecución, la latencia y el costo de ejecución.
- Scheduling de escalabilidad, descarga e ingresos.
- Métodos de scheduling inteligente.

Una característica de FaaS es la composición de funciones: una función puede llamar (secuencialmente) a otra función, que, a su vez, puede llamar a otra función en una cadena de invocaciones [23]. Desde el punto de vista de la infraestructura, una composición de funciones es menos opaco que una máquina virtual o un contenedor, conociendo la cadena de funciones el scheduler puede preparar el entorno de ejecución con anticipación [24].

El proceso de scheduling aplicado a Serverless, resulta un gran desafío debido a varios factores, como el comportamiento dinámico, los recursos heterogéneos, la carga de trabajo que varían en volumen, las características distintivas de la arquitectura y las variaciones en la cantidad de solicitudes.

.Objetivos

Los objetivos del grupo de investigación en esta línea de conocimiento son los siguientes:

- Analizar los diferentes algoritmos de scheduling.
- Establecer cuales de las estrategias de scheduling pueden ser aplicadas en entorno de Serverless Computing
- Proponer y evaluar estrategias para lograr eficiencia y mejoras en los algoritmos seleccionados.
- Evaluar y medir el comportamiento de los algoritmos en entornos serverless.

Los análisis se desarrollarán con independencia de la plataforma.

Formación de Recursos Humanos

El equipo de trabajo está compuesto por los investigadores que figuran en este trabajo y dos alumnos de grado, de la Universidad Nacional de San Juan. Además, el proyecto marco donde se está desarrollando esta propuesta ha establecido vínculos con investigadores de la Nacional de San Luis, de la Universidad Champagnat y de la Universidad Nacional de Salta, y dos alumnos de grado más.
Se está desarrollando una tesis doctoral sobre paralelismo híbrido y Big Data, una tesis de maestría en áreas afines y dos tesinas de grado en el área de Serverless computing, Concurrencia y Computación distribuida.
Por otro lado se espera aumentar el número de publicaciones, y además también se prevé la divulgación de varios temas investigados por medio de cursos de postgrado y actualización o publicaciones de divulgación y asesoramiento a empresas y otras instituciones públicas y privadas.

Referencias

[1] Venkata R. Vutti. (2024). Accelerating Digital Transformation Through Serverless Architecture. International Journal of Research in Computer Applications and Information Technology (IJRCAIT), 7(2), 2297.

[2] Adzic, G., Chatley, R. (2017). Serverless computing: economic and architectural impact. In Proceedings of the 2017 11th Joint Meeting on Foundations of Software Engineering (pp. 884-889). ACM.

[3] https:// aws.amazon.com/es/lambda/

[4] Ghorbian M, Ghobaei-Arani M, and Esmaeili L (2024). A survey on the scheduling mechanisms in serverless computing: a

taxonomy, challenges, and trends Cluster Comput. 2024 27 5571.

[5] K. Kaffes, N. J. Yadwadkar, and C. Kozyrakis (2019), Centralized Coregranular Scheduling for Serverless Functions, SoCC,

[6] Serverless Computing. (2024) Guest Editors' Introduction. IEEE Internet Computing.

[7] .Cloud computing: A complete guide. IBM. https://www.ibm.com/cloud/learn/cloud-computing-gbl

[8] M. Armbrust, et al., Above the clouds (2009). A Berkeley view of cloud computing, In: Tech. Rep. No. UCB/EECS-2009-28,

[9] Keith D. Foote: (2021). A Brief History of Cloud Computing. https://www.dataversity.net/brief-history-cloud-computing/

[10] A research and markets report (2024). https://omdia.tech.informa.com/pr/2024/jun/omdia-serverless-computing-valued-at-19-billiondollars-is-the-fastest-growing-cloud-service.

[11] Mohanty, S. K., Premsankar, G., & Di Francesco, M. (2018). An Evaluation of Open Source Serverless Computing Frameworks. In CloudCom (pp. 115-120).

[12] Gottlieb, N. (2016). State of the Serverless Community Survey Results. https: //serverless.com/blog/state-of-serverless-community/.

[13] S. Chippagiri (2025). The Rise of Serverless Computing: A Systematic Review of Challenges and Solutions with Optimization Strategies. The Review of Contemporary Scientific and Academic Studies. Vol. 5 | Issue No. 01 | 25

[14] M. Ghorbian , M. Ghobaei-Arani. Serverless Computing: Architecture, Concepts, and Applications. https://arxiv.org/abs/2501.09831.

[15] Scheduling. Computer Science Wiki. https://computersciencewiki.org/index.php/Scheduling.

[16] Rajni Aron, Ajith Abraham (2022). Resource scheduling methods for cloud computing environment: The role of meta-heuristics and artificial intelligence, Engineering Applications of Artificial Intelligence, Volume 116, 2022-

[17] A. Mampage, S. Karunasekera, R. Buyya (2023). Deep reinforcement learning for application scheduling in resource-constrained, multi-tenant serverless computing environments. Future Generation Computer Systems Vol. 143, June 2023, Pages 277-292

[18] Thomas Rausch et al. (2021). Optimized container scheduling for data-intensive serverless edge. Computing. Future Generation Computer Systems 114 259–271.

[19] Anupama Mampage et al. Deep reinforcement learning for application scheduling in resource-constrained, multi-tenant serverless computing environments. Future Generation Computer Systems 143 (2023) 277–292

[20] Sebastián Risco, et al. Rescheduling serverless workloads across the cloud-to-edge continuum. Future Generation Computer Systems 153 (2024) 457–466

[21] P. Zuk, , K. Rzadca (2022). Reducing response latency of composite functions-as-a-service through scheduling. Journal of Parallel and Distributed Computing 167 18–30.

[22] A. Ebrahem, B. Mahmmood (2023). A Review: Scheduling Methods in Serverless Edge Computing. International Research Journal of Innovations in Engineering and Technology (IRJIET). Volume 7, Issue 2, pp 36-41.

[23] Nelson Rodríguez et al. (2021). Análisis de ejecución múltiple de Funciones Serverless en AWS. XXVII CACIC.

[24] P. Zuk, , K. Rzadca.(2020). Scheduling Methods to Reduce Response Latency of

Function as a Service. IEEE 32nd International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD)

· RCCI ·

Redes de Cooperación Científica Internacionales

Acciones de colaboración en el Consorcio de Cloud Computing, Big Data y Emerging Topics durante 2024 (CCC-BD&ET)

III-LIDI – Instituto de Investigación en Informática LIDI (UNLP – Argentina)
LISSI–Laboratorio de Investigación y Desarrollo en Ingeniería de Software y Sistemas de Información (UNS – Argentina)
VyGLab – Laboratorio de Investigación y Desarrollo en Visualización y Computación Gráfica (UNS– Argentina)
LIDIC – Laboratorio de Investigación y Desarrollo en Inteligencia Computacional (UNSL – Argentina)
HPC4EAS – High Performance Computing for Efficient Applications and Simulation (UAB – España)
SMILe – Soft Management of Internet and Learning (Universidad de Castilla-La Mancha – España)
ArTeCS – Group of Architecture and Technology of Computing Systems (UCM – España)
LITRP – Laboratorio de Investigaciones Tecnológicas en Reconocimiento de Patrones (UCM – Chile)
LCG – Laboratorio de Computación Gráfica (UNSL – Argentina)
ITIC – Instituto para las Tecnologías de la Información y las Comunicaciones (UNCu – Argentina)
DisCo – Grupo de I+D+I en Computación Distribuida (Universidad de Zaragoza – España)
& Investigadores Asociados al CCC- BD&ET

Resumen

El Consorcio de I+D+i en *Cloud Computing, Big Data & Emerging Topics* (CCC-BD&ET) [1] es una iniciativa para fomentar y formalizar la colaboración existente entre grupos de investigación de varias universidades en temáticas vinculadas a Cloud Computing, al Análisis Masivo de Datos y a Tópicos Emergentes, como las tecnologías 4.0 o los gemelos digitales aplicados a distintos ámbitos, entre otros. Estas temáticas, y su integración, han adquirido creciente importancia por su aplicación en dominios de alto impacto como las ciudades inteligentes, la transformación digital, los sistemas de E-health y los basados en tecnologías tales como IoT, IIoT y Blockchain.

Los integrantes del consorcio, provenientes mayoritariamente de Argentina, Chile y España han tenido, a lo largo de los años, diversas experiencias de trabajo conjunto que fueron consolidadas a partir de la organización y realización de las *Jornadas de Cloud Computing-Big Data & Emerging Topics* (JCC-BD&ET) llevadas a cabo en la Universidad Nacional de La Plata (Argentina). La constitución de este Consorcio, reafirma y formaliza estas líneas de colaboración proponiendo acciones de cooperación académica vinculadas con la formación de recursos humanos, la formulación y ejecución de proyectos conjuntos, y la vinculación con empresas y organismos relacionados con la industria informática, entre otras.

Palabras clave: Cloud Computing, Big Data, Colaboración interinstitucional, Cooperación internacional.

Contexto

El Consorcio de I+D+i en *Cloud Computing, Big Data & Emerging Topics* (CCC-BD&ET) resulta de la cooperación, llevada a cabo durante varios años, entre grupos de investigación, desarrollo e innovación de Universidades vinculadas con las realización anual de las *Jornadas de Cloud Computing-Big Data & Emerging Topics* (JCC-BD&ET), organizadas por la Universidad Nacional de La Plata (UNLP), en Argentina.

I. Introducción

Las Jornadas de Cloud Computing, Big Data & Emerging Topics (JCC-BD&ET) son un encuentro anual de intercambio de ideas, proyectos, resultados científicos y aplicaciones concretas en diferentes áreas relacionadas con Cloud Computing, Inteligencia de Datos, Big Data y Tecnologías Emergentes.

Las Jornadas integran ponencias científicas con experiencias de desarrollos y aplicaciones, fomentando la interacción entre la academia y los sectores productivos/industriales, en el área temática de Cloud Computing, Big Data y Tecnologías Emergentes.

En el marco de las Jornadas también se desarrollan Conferencias, Paneles y Cursos de Posgrado específicos.

Las JCC-BD&ET son organizadas por el Instituto de Investigación en Informática LIDI (III-LIDI) y la Secretaría de Posgrado de la Facultad de Informática de la UNLP en colaboración con Universidades de Argentina y del exterior. Cuentan con diferentes auspicios de organismos de Ciencia y Tecnología de Argentina y de sectores representativos de la industria del Software de Argentina. Se iniciaron en 2013 con foco en los temas de Cloud Computing. A partir del 2015, pasaron a ser las Jornadas de Cloud Computing & Big Data y, en el año 2020, se constituyeron en JCC-BD&ET.

A partir de estos encuentros fueron surgiendo distintas actividades en colaboración, que no sólo se formalizaron mediante acuerdos entre los distintos actores. Los resultados de las colaboraciones se encuentran reflejados en las publicaciones detalladas en las páginas mencionadas en las referencias.

Objetivos

Como corolario de las distintas actividades de colaboración realizadas por los participantes a las JCC-BD&ET es que surge la propuesta de conformar un Consorcio de I+D+i en Cloud Computing, Big Data & Emerging Topics que permita afianzar y proyectar a futuro las relaciones existentes [1]. Bajo el amparo de las JCC-BD&ET nuestro objetivo es seguir constituyendo un foco anual de encuentro, independientemente del trabajo académico y científico que se realice durante el año para la concreción de los objetivos del Consorcio, el fortalecimiento de la cooperación, la formación de recursos humanos y la contribución a la comunicación científica en el área.

Cooperaciones actuales entre Universidades del Consorcio

En el marco del consorcio, sus integrantes han establecido colaboraciones plasmadas en los siguientes proyectos que han sido formalizados en diferentes convocatorias:

a) Proyecto "Computación de Alto Desempeño: Arquitecturas, Algoritmos, Métricas de Rendimiento y Aplicaciones en HPC, Big Data, Robótica, Señales y Tiempo Real." UNP 2023-2026, con la colaboración de UAB, UCM, UNRN, UNGS, UNCOMA y UNDEC.
b) Proyecto "Inteligencia de Datos. Técnicas y Modelos de Machine Learning" UNLP, 2023-2026, con la colaboración de UCLM y URV.
c) Proyecto "Diseño, desarrollo y evaluación de sistemas en escenarios híbridos para áreas clave de la sociedad actual: educación, ciudades inteligentes y gobernanza digital" UNLP 2023-2026, con la colaboración de UNS, UCLM y UNIZAR.
d) Proyecto "Computación de Altas Prestaciones Eficiente y Segura para Aplicaciones de Servicios de Salud Inteligentes" UAB, con la colaboración de UNLP y UNSL.
e) Proyecto UNI-UEAR "Alianza Universitaria Argentina Europea para la Transformación Digital". Consorcio internacional con participación de UNLP y UNS. (desde 2022)
f) Proyecto REACT4CITIES "Resilience, Empowerment, and AI-powered Capacity-building for Transformation in Cities and Communities". Consorcio internacional con participación de UNLP y UNS (en evaluación)
g) Programa GEMINAE - UNITA. "Cooperación de universidades asociadas en Geminae con los integrantes de UNITA para las mejores prácticas educativas." Consorcio internacional con participación de UNIZAR y UNLP (desde 2023)
h) Proyecto TED: "Ecosistema tecnológico para el reconocimiento del estado de ánimo de los pacientes durante el proceso de rehabilitación cardíaca. Acrónimo del

proyecto: TEMOR" UNIZAR, con la participación de UNLP.

i) Proyecto "Experiencias Lúdicas con Agentes Sociales Interactivos y Robots: Aprendizaje Social y Comunicación Intergeneracional (PLEISAR)" UNIZAR, con la participación de UNLP.
j) Proyecto "Tecnologías avanzadas aplicadas al procesamiento de datos masivos" UNSL con la colaboración de UAB.

Resultados Obtenidos

Los resultados esperados están relacionados con el fortalecimiento de las capacidades de los grupos de investigación del Consorcio, en términos de sus RRHH, redes de trabajo, acceso a recursos organizacionales y mejoramiento de la calidad de los resultados producidos, producto de la colaboración interdisciplinaria e inter-universitaria promovida por el Consorcio. Los resultados mas relevantes se mencionan en las siguientes subsecciones.

Formación de RRHH

Los integrantes del CCC-BD&ET están abocados a la formación de recursos humanos tanto a nivel de grado como de posgrado, en temáticas afines al Consorcio, en particular en muchos casos se realiza colaboración entre los distintos grupos, a través de participación en los comités académicos de carreras, dictado de cursos, dirección y evaluación de tesis.

Por otro lado, fruto de las interacciones entre investigadores en diferentes temas, han surgido cursos e incluso nuevas carreras de Postgrado que tienen intervención de grupos de I+D+I del Consorcio.

Entre estas carreras y cursos se pueden mencionar:

a) Maestría "Cómputo de Altas Prestaciones" (UNLP); Especialización "Cómputo de Altas Prestaciones y Tecnología GRID" (UNLP) y la Especialización "Cómputo de Altas Prestaciones" (UNLP) se dictan con participación de UNS, UNSL, UAB y UCM.
b) Maestría "Gestión y tecnología de ciudades inteligentes" (UNS-UNLP).
c) Maestría "Tecnología Informática Aplicada en Educación" (UNLP) y la Especialización "Tecnología Informática Aplicada en Educación" (UNLP) se dictan con participación de UAB y UNIZAR.
d) Maestría "Inteligencia de datos orientada a Big Data" y la Especialización "Inteligencia de datos orientada a Big Data" (UNLP) se dictan con participación de UCLM y URV.
e) Especialización "Computación Gráfica, Imágenes y Visión por Computadora" (UNLP) se dicta con participación de UNS y UNSL.
f) Curso de postgrado "Gemelos digitales: aplicaciones industriales y sociales" (UNSL-UNLP-UAB). 2024.

En el mismo sentido, se resaltan las colaboraciones establecidas para la (co-) dirección de estudiantes de posgrado:

a) Dirección Tesis Doctoral "Estudio de Viabilidad de SYCL como Modelo de Programación Unificado para Sistemas Heterogéneos Basados en GPUs en Bioinformática" M. Costanzo.(UNLP-UCM) Egresado 2024
b) Dirección Tesis Doctoral "Desarrollo de un modelo de elicitación de emociones a partir de las características de la música. Generación de un sistema recomendador" Y. Ospitia Medina.(UNLP-UNIZAR) Egresado 2024
c) Dirección Tesis Doctoral "Propuesta de un modelo de gamificación para entorno virtuales de enseñanza y aprendizaje en instituciones de educación superior y su impacto en la presencia social, cognitiva y docente en dichos entornos" G. R. Vera Mora.(UNLP-UNIZAR) Egresada 2024
d) Dirección Tesis Doctoral "Modelización y gestión del consumo energético en un sistema de altas prestaciones con tolerancia de fallos" M. Morán.(UNLP-UAB-UNCOMA). En evaluación.
e) Dirección Tesis Doctoral "Caracterización de aplicaciones con comportamiento irregular para predecir su rendimiento, basado en la filosofía PAS2P" F.

Tirado.(UAB-UCM Chile). En curso

f) Dirección Tesis Doctoral “Diseño de un modelo computacional basado en agentes para simular servicios de telemedicina” E. Solís Acosta.(UNLP-UAB). En curso.
g) Dirección Tesis Doctoral “Modelizacion y Simulacion basada en Agentes Aplicada a la Arquitecturas de Entrada/Salida de los computadores Paralelos” D. Encinas.(UNLP-UAB). En curso.
h) Dirección Tesis Doctoral “Desarrollo de un modelo de relaciones entre videos educativos, perfiles de estudiantes y emociones” G. Astudillo. (UNLP-UNIZAR) En curso
i) Dirección de Tesis Doctoral “Modelo de predicción orientado a la generación de residuos electrónicos con un enfoque de Ciudades Inteligentes Sostenibles desde la perspectiva de la Inteligencia Ambiental” Facuy Delgado Jussen Paul. (UNLP-UNS) En curso
j) Dirección de Tesis Doctoral “Modelo digital para la reducción del impacto ambiental de las actividades agrícolas utilizando variables climatológicas en tiempo real aplicado al cordón frutihortícola del gran La Plata” Leonardo Navarria (UNLP-UNS) En curso
k) Dirección Tesis de Maestría en Tecnología Informática Aplicada en Educación “Videojuego como método de evaluación sigilosa para analizar rasgos de personalidad vinculados al rendimiento académico” P. Ferreyra (UNLP-UNIZAR) Egresado 2024

Asimismo los grupos tienen reuniones de intercambio de posibilidades de apoyo en actividades de postgrado para alumnos de Maestría y Doctorado.
Esto incluye financiamiento para estadías de investigación y dictado de cursos abiertos a las Universidades del Consorcio, así como la definición de Ciclos y Estadías Postdoctorales.

Producción científica conjunta (2024-3/2025)

a) “Acciones de colaboración en el Consorcio de Cloud Computing, Big Data y Emerging Topics (CCC-BD&ET)” III-LIDI (UNLP), LISSI (UNS), VyGLab (UNS), LIDIC (UNSL), HPC4EAS (UAB), SMILe (UCLM), ArTeCS (UCM, España), LITRP (UCM, Chile), LCG (UNSL), ITIC (UNCu), DisCo (UZ) & Investigadores Asociados al CCC-BD&ET. Libro de Actas XXVI Workshop de Investigadores en Ciencias de la Computación (WICC 2024). ISBN: 978-987-8352-57-2. Págs. 699-705. https://wicc2024.unp.edu.ar/images/libros/LibroDeActasWICC2024.pdf
b) “PlateUNLP: An Open-Source Solution for Automating the Digitization of Historical Astronomical Plates”. Santiago Andres Ponte Ahón, Yael Aidelman, Juan Martín Seery, Facundo Quiroga, Franco Ronchetti, Waldo Hasperué, M. Iannuzzi, R. Peralta, M. López, Aurelio F. Bariviera, Lydia Cidale, Roberto Gamen. En: GCH 2024 - Eurographics Workshop on Graphics and Cultural Heritage, ISBN 978-3-03868-248-6, The Eurographics Association, 2024
c) “FGR-Net: Interpretable fundus image gradeability classification based on deep reconstruction learning”. Saif Khalid, Hatem A. Rashwan, Saddam Abdulwahab, Mohamed Abdel-Nasser, Facundo Manuel Quiroga, Domenec Puig. Expert Systems with Applications (ISSN 0957-4174). Vol. 238, Part C. (121644). Marzo 2024. https://doi.org/10.1016/j.eswa.2023.121644
d) “Model for Designing Gamified Experiences Mediated by a Virtual Teaching and Learning Environment”. Glenda Vera-Mora, Cecilia V. Sanz, Teresa Coma-Roselló, Sandra Baldassarri. Education Sciences (ISSN 2227-7102). Vol. 14, Núm. 8. Págs. 907. Agosto 2024. https://doi.org/10.3390/educsci14080907
e) “Developing educational serious games as a proposal for university training in the area of Child Computer Interaction”. Cecilia Verónica Sanz, Verónica Artola, Sandra Baldassarri. International Journal of Child-Computer Interaction (ISSN 2212-8689). Vol. 41. Págs. 100672. Septiembre 2024.

https://doi.org/10.1016/j.ijcci.2024.100672

f) “VISTA: vision improvement via split and reconstruct deep neural network for fundus image quality assessment”. Saif Khalid, Saddam Abdulwahab, Oscar Agustín Stanchi, Facundo Manuel Quiroga, Franco Ronchetti, Domenec Puig, Hatem A. Rashwan. Neural Computing and Applications (ISSN 1433-3058) Octubre 2024. https://doi.org/10.1007/s00521-024-10174-6

g) “Análisis de desempeño de Servicio de Urgencia Hospitalario con llegada creciente de pacientes”. Mariela Rodríguez, Francesc Boixader, Francisco Epelde, Eva Bruballa, Armando De Giusti, Álvaro Wong, Dolores Rexachs, Emilio Luque. Libro de Actas del 30° Congreso Argentino de Ciencias de la Computación (CACIC 2024). Págs. 176-185. ISBN: 978-950-34-2428-5. Octubre 2024. https://sedici.unlp.edu.ar/handle/10915/172755

h) “Evaluación del tiempo crítico para realizar acciones preventivas en el Servicio de Urgencias Hospitalarias”. Mercedes Planas, Eva Bruballa, Emilio Luque, Dolores Rexachs, Francisco Epelde, Armando De Giusti. Libro de Actas del 30° Congreso Argentino de Ciencias de la Computación (CACIC 2024). Págs. 218-227. ISBN: 978-950-34-2428-5. Octubre 2024. https://sedici.unlp.edu.ar/handle/10915/172755

i) “Introducción al modelo digital para la reducción del impacto ambiental de las actividades agrícolas”. Leonardo Navarria, Ariel Pasini, Elsa Estevez. Libro de Actas del 30° Congreso Argentino de Ciencias de la Computación (CACIC 2024). Págs. 1363-1373. ISBN: 978-950-34-2428-5. Octubre 2024. https://sedici.unlp.edu.ar/handle/10915/172755

j) “Modelo digital para la reducción del impacto ambiental de las actividades agrícolas en tiempo real. Revisión de Literatura”. Leonardo Navarria, Ariel Pasini, Elsa Estevez. Libro de Actas del 30° Congreso Argentino de Ciencias de la Computación (CACIC 2024). Págs. 1384-1395. ISBN: 978-950-34-2428-5. Octubre 2024. https://sedici.unlp.edu.ar/handle/10915/172755

k) “Modelo predictivo aplicando inteligencia ambiental orientado a la generación de residuos electrónicos con un enfoque de Ciudades inteligentes sostenible. Revisión de Literatura”. Facuy Jussen, Ariel Pasini, Elsa Estevez, César Moran. Libro de Actas del 30° Congreso Argentino de Ciencias de la Computación (CACIC 2024). Págs. 1396-1405. ISBN: 978-950-34-2428-5. Octubre 2024. https://sedici.unlp.edu.ar/handle/10915/172755

l) “Quantitative Evaluation of White & Black Box Interpretability Methods for Image Classification”. Oscar Stanchi, Franco Ronchetti, Pedro Dal Bianco, Gastón Ríos, Waldo Hasperué, Domenec Puig Valls, Hatem Rashwan, Facundo Quiroga. Libro de Actas del 30° Congreso Argentino de Ciencias de la Computación (CACIC 2024). Págs. 125-134. ISBN: 978-950-34-2428-5. Octubre 2024. https://sedici.unlp.edu.ar/handle/10915/172755

m) “Exploring energy saving opportunities in fault tolerant HPC systems”. Marina Morán, Javier Aldo Balladini, Dolores Rexachs, Enzo Rucci:J. Parallel Distributed Comput. 185: 104797 (2024)

n) “Revisión sistemática sobre videojuegos como método de evaluación sigilosa para analizar rasgos de personalidad”. Paula Ferreyra, Teresa Coma-Roselló, Cecilia Sanz. Libro de Actas XIX Congreso de Tecnología en Educación y Educación en Tecnología (TE&ET 2024). Págs. 37-47. ISBN: 978-987-3977-95-4. Junio 2024. https://sedici.unlp.edu.ar/handle/10915/171424

o) “Agent Based Simulation as an Efficient Way for HPC I/O System Tuning”. Diego Encinas, Marcelo Naiouf, Armando De Giusti, Román Bond, Sandra Mendez,

Dolores Rexachs, Emilio Luque. Computational Science – ICCS 2024. ICCS 2024. Lecture Notes in Computer Science. Vol 14834. ISBN 978-3-031-63759-9. Págs. 129-136. Cham, 2024. https://doi.org/10.1007/978-3-031-63759-9_16

p) "Arquitecturas Multiprocesador en HPC: Aspectos de Portabilidad, Resiliencia y Eficiencia Energética". Armando De Giusti, Marcelo Naiouf, Fernando G. Tinetti, Horacio Villagarcía Wanza, Franco Chichizola, Laura De Giusti, Enzo Rucci, Adrián Pousa, Victoria Sanz, Diego Montezanti, Diego Encinas, Ismael Rodriguez, Sebastián Rodriguez Eguren, Leandro Libutti, Juan Manuel Paniego, Martin Pi Puig, César Estrebou, Javier Balladini, Manuel Costanzo. Libro de Actas XXVI Workshop de Investigadores en Ciencias de la Computación (WICC 2024). ISBN: 978-987-8352-57-2. Págs. 630-636. https://wicc2024.unp.edu.ar/images/libros/LibroDeActasWICC2024.pdf

q) "Cómputo paralelo y distribuido: eficiencia energética y aplicaciones en salud". Javier Balladini, Marina Morán, Claudio Zanellato, Rodrigo Cañibano, Manuel Latorre, Cristina Orlandi, Enzo Rucci, Armando De Giusti, Remo Suppi, Dolores Rexachs, Emilio Luque, Emmanuel Frati. Libro de Actas XXVI Workshop de Investigadores en Ciencias de la Computación (WICC 2024). ISBN: 978-987-8352-57-2. Págs. 684-687. https://wicc2024.unp.edu.ar/images/libros/LibroDeActasWICC2024.pdf

r) "Deep Learning Para Visión por Computadora". Facundo Quiroga, Franco Ronchetti, Gastón Ríos, Pedro Dal Bianco, Santiago Ponte Ahon, Oscar Stanchi, Juan Seery, Tatiana Badaracco, Federico Rabinovich, Javier Vázquez, Aurelio Fernandez Bariviera, Saif Khalid, Hatem Rashwan, Domenee Puig Valls, Yael Aidelman, Roberto Gamen, Laura Lanzarini, Waldo Hasperué. Libro de Actas XXVI Workshop de Investigadores en Ciencias de la Computación (WICC 2024). ISBN: 978-987-8352-57-2. Págs. 88-92. https://wicc2024.unp.edu.ar/images/libros/LibroDeActasWICC2024.pdf

s) "Diseño y desarrollo de experiencias educativas con juegos serios y otras tecnologías digitales". Cecilia Sanz, Verónica Artola, Natalí Salazar Mesía, Luciano Iglesias, Abril Buffarini, Bárbara Ibañez, Laureano Bruno, Matías Zeballos, Aldana Del Gener, Gustavo Javier Astudillo, Sandra Baldassarri. Libro de Actas XXVI Workshop de Investigadores en Ciencias de la Computación (WICC 2024). ISBN: 978-987-8352-57-2. Págs. 873-879. https://wicc2024.unp.edu.ar/images/libros/LibroDeActasWICC2024.pdf

t) "Gemelos Digitales: Aplicación en la Industria de Procesos, en la Gestión Sanitaria y en el área Agrícola". Federico Walas Mateo, Dolores Rexachs, Merce Planas, Eva Bruballa, Francisco Epelde, Alvaro Wong, Emilio Luque, Joel Acosta, Sonia Fretes, Ayelen Elisabet Cayuqueo, Gustavo Guitera, Claudia Russo, Leonardo Esnaola, Hugo Ramón, Adrian Jaszczysyzn, Laura De Giusti, Marcelo Naiouf, Santiago Medina, Armando De Giusti. Libro de Actas XXVI Workshop de Investigadores en Ciencias de la Computación (WICC 2024). ISBN: 978-987-8352-57-2. Págs. 653-656. https://wicc2024.unp.edu.ar/images/libros/LibroDeActasWICC2024.pdf

u) "Investigación y desarrollo de Tecnologías Digitales para Educación". Cecilia Sanz, Verónica Artola, Alejandro Gonzalez, Maria Alejandra Zangara, Luciano Iglesias, Eduardo Ibañez, Lucía Violini, Adriana Fachal, Aldana Del Gener, María José Abásolo, Cristina Manresa Yee, María Paula Dieser, Teresa Coma-Roselló, Lorenzo Handula, Sabrina Lombardo, Patricia Pesado. Libro de Actas XXVI Workshop de Investigadores en Ciencias de la Computación (WICC 2024). ISBN: 978-987-8352-57-2. Págs. 880-886. https://wicc2024.unp.edu.ar/images/libros/LibroDeActasWICC2024.pdf

v) "Modelado de un Servicio de Emergencia Resiliente". Mariela Rodriguez, Francesc Boixader, Francisco Epelde, Eva Bruballa, Armando De Giusti, Alvaro Wong, Dolores Rexachs, Emilio Luque. Libro de Actas XXVI Workshop de Investigadores en Ciencias de la Computación (WICC 2024). ISBN: 978-987-8352-57-2. Págs. 157-162. https://wicc2024.unp.edu.ar/images/libros/LibroDeActasWICC2024.pdf

w) "Paralelización de algoritmos y evaluación de rendimiento en plataformas de cómputo de altas prestaciones. Aplicaciones". Marcelo Naiouf, Armando De Giusti, Franco Chichizola, Victoria Sanz, Adrián Pousa, Enzo Rucci, María José Basgall, Mariano Sanchez, Manuel Costanzo, Fernando Emmanuel Frati, Adriana Angélica Gaudiani, Sergio Leandro Calderón, Laura De Giusti. Libro de Actas XXVI Workshop de Investigadores en Ciencias de la Computación (WICC 2024). ISBN: 978-987-8352-57-2. Págs. 647-652. https://wicc2024.unp.edu.ar/images/libros/LibroDeActasWICC2024.pdf

x) "Procesamiento inteligente de grandes volúmenes de datos". César Estrebou, Genaro Camele, Enzo Rucci, Franco Ronchetti, Sebastián Rodriguez Eguren, David Castillo, Gary Reyes Zambrano, Laura Lanzarini Aurelio F. Bariviera, Waldo Hasperué. Libro de Actas XXVI Workshop de Investigadores en Ciencias de la Computación (WICC 2024). ISBN: 978-987-8352-57-2. Págs. 203-207. https://wicc2024.unp.edu.ar/images/libros/LibroDeActasWICC2024.pdf

Integrantes del Consorcio

El Consorcio está integrado tanto por grupos de investigación como por investigadores asociados. A continuación se detallan los mismos.

Grupos que conforman el CCC- BD&ET

III-LIDI – Instituto de Investigación e Informática LIDI (Universidad Nacional de La Plata – Argentina) [6]
Directora: *Lic. Patricia Pesado*

LISSI – Laboratorio de Investigación y Desarrollo en Ingeniería de Software y Sistemas de Información (Universidad Nacional del Sur – Argentina) [7]
Director: *Dr. Pablo Fillottrani*

VyGLab – Laboratorio de Investigación y Desarrollo en Visualización y Computación Gráfica (Universidad Nacional del Sur – Argentina) [8]
Directora: *Dra. Silvia Castro*

LIDIC – Laboratorio de Investigación y Desarrollo en Inteligencia Computacional (Universidad Nacional de San Luis – Argentina) [9]
Director: *Dr. Marcelo Errecalde*

HPC4EAS – High Performance Computing for Efficient Applications and Simulation (Universidad Autónoma de Barcelona – España) [10]
Director: *Dr. Emilio Luque*

SMILe – Soft Management of Internet and Learning (Universidad de Castilla-La Mancha – España) [11]
Director: *Dr. José A. Olivas Varela*

ArTeCS – Group of Architecture and Technology of Computing Systems (Universidad Complutense de Madrid – España) [12]
Director: *Dr. Francisco Tirado*

LITRP – Laboratorio de Investigaciones Tecnológicas en Reconocimiento de Patrones (Universidad Católica de Maule – Chile) [13]
Representante para el Consorcio: *Dr. Ricardo Barrientos*

LCG – Laboratorio de Computación Gráfica (Universidad Nacional de San Luis–Argentina) [14]
Director: *Mg. Roberto Guerrero*

ITIC – Instituto para las Tecnologías de la Información y las Comunicaciones (Universidad Nacional de Cuyo – Argentina)
[15].
Director: *Dr. Carlos García Garino*

DisCo – Grupo de I+D+I en Computación Distribuida (Universidad de Zaragoza – España) [16].

Coordinador: *Dr. Pedro Javier Álvarez Pérez-Aradros*

Investigadores Asociados al CCC- BD&ET

Dr. Aurelio Fernández (Universidad Rovira i Virgili – España).
Dr. Emmanuel Frati (Universidad Nacional de Chilecito – Argentina).
Dr. Javier Balladini (Universidad Nacional de Comahue – Argentina).
Dra. Adriana Gaudiani (Universidad Nacional de General Sarmiento – Argentina).
Dra. Mónica Denham (Universidad Nacional de Río Negro – Argentina).
Lic. Nelson Rodríguez (Universidad Nacional de San Juan – Argentina).

Referencias

[1] Consorcio de I+D+I en Cloud Computing, Big Data & Emerging Topics (CCC-BD&ET), https://jcc.info.unlp.edu.ar/consorcio-de-idi-en-cloud-computing-big-data-emerging-topics/.
[2] Consorcio para la colaboración en I+D+I en temas de Cloud Computing, Big Data y Emerging Topics (CCC-BD&ET)- WICC 2021. 978-987-24611-3-3; 978-987-24611-4-0. Páginas: 778-783.
[3] "Consorcio para la colaboración en I+D+I en Temas de Cloud Computing, Big Data y Emerging Topics (CCC-BD&ET) Proyecto Integrador: "Transformación Digital en la incorporación de la Resiliencia como un Key Performance Indicator de Prestaciones Sociales (KPIS)". III-LIDI (UNLP) LISSI (UNS), VyGLab (UNS), LIDIC (UNSL), HPC4EAS (UAB), SMILe (UCLM), ArTeCS (UCM, España), LITRP (UCM, Chile), LCG (UNSL), ITIC (UNCu), DisCo (UZ) & Investigadores Asociados al CCC- BD&ET. XXIV WICC 2022. ISBN 978-987-48222-3-9. Págs. 574-578.
[4] "Avances del Consorcio para la colaboración en I+D+I en Temas de Cloud Computing, Big Data y Emerging Topics (CCC-BD&ET)" III-LIDI (UNLP), LISSI (UNS), VyGLab (UNS), LIDIC (UNSL), HPC4EAS (UAB), SMILe (UCLM), ArTeCS (UCM, España), LITRP (UCM, Chile), LCG (UNSL), ITIC (UNCu), DisCo (UZ) & Investigadores Asociados al CCC-BD&ET. WICC 2023. ISBN 978-987-3724-66-4.
[5] "Acciones de colaboración en el Consorcio de Cloud Computing, Big Data y Emerging Topics (CCC-BD&ET)" III-LIDI (UNLP), LISSI (UNS), VyGLab (UNS), LIDIC (UNSL), HPC4EAS (UAB), SMILe (UCLM), ArTeCS (UCM, España), LITRP (UCM, Chile), LCG (UNSL), ITIC (UNCu), DisCo (UZ) & Investigadores Asociados al CCC-BD&ET. WICC 2024. ISBN: 978-987-8352-57-2. Págs. 699-705. https://wicc2024.unp.edu.ar/images/libros/LibroDeActasWICC2024.pdf
[6]http://weblidi.info.unlp.edu.ar
[7]https://lissi.cs.uns.edu.ar
[8]http://vyglab.cs.uns.edu.ar
[9]http://lidic.unsl.edu.ar
[10] https://grupsderecerca.uab.cat/hpc4eas/
[11]http://smile.esi.uclm.es
[12]https://artecs.dacya.ucm.es
[13]http://www.litrp.cl
[14]http://www.lcg.unsl.edu.ar/
[15] http://itic.uncu.edu.ar/
[16] http://webdiis.unizar.es/DISCO/

Participación en el Programa UNITA GEMINAE

Instituto de Investigación en Informática LIDI (III-LIDI)
Facultad de Informática – Universidad Nacional de La Plata
50 y 120 La Plata Buenos Aires
Centro Asociado CIC - 526 e/ 10 y 11 La Plata Buenos Aires

Directora III-LIDI: Patricia Pesado [0000-0003-0000-3482]
Coordinadora en GEMINAE por el III-LIDI: Cecilia Sanz (1) [0000-0002-9471-0008]
(ppesado, csanz) @lidi.info.unlp.edu.ar

Resumen

Geminae es un programa de la Alianza Universitaria UNITA que busca crear una red de universidades para compartir buenas prácticas. El objetivo es mejorar la movilidad de los estudiantes entre las instituciones y fomentar que se establezcan en las regiones de UNITA. La red de universidades se basa en la calidad de las instituciones y en la cooperación activa entre universidades fuera de la Unión Europea y los socios de UNITA.

Palabras clave: Buenas prácticas – innovación en educación

Contexto

Desde el año 2023 el Instituto de Investigación en Informática LIDI (III-LIDI) participa en el programa UNITA Geminae.

Introducción

El Programa GEMINAE UNITA se orienta a generar una red de universidades para lograr el intercambio de buenas prácticas desarrolladas en los ámbitos de formación, investigación, innovación y vinculación con la sociedad.
El programa se basa en la cooperación activa entre universidades fuera de la Unión Europea: UNLP, ITBA, UNCuyo (Argentina), FBS, UB, UEC, UEOP, UFB, UFMG, UFV, UFRJ, URB (Brasil), UQTR, US (Canadá), ULH (Cuba), UCC, UA, UR, UNC (Colombia), UASLP, UM (México), UASD (Rep. Dominicana); UB (Argelia), UAC (Benín), 2IE (Burkina Faso), UCV (Cabo Verde), UA (Madagascar), UH2C (Marruecos), UPM (Mozambique), UNTL (Timor-Leste)) y los socios de UNITA (Agencia Universitaria de la Francofonía – AUF, Organización de Estados Iberoamericanos – OEI, Asociación de Universidades de Lengua Portuguesa – AULP.

Objetivos

Su objetivo es proporcionar competencias y habilidades que permitan a los estudiantes convertirse en futuros actores de economías sostenibles, basadas en la inclusión y la igualdad, teniendo un impacto local, regional y global.

En particular:

- Intercambiar buenas prácticas en formación, investigación, innovación y vinculación con la sociedad
- Facilitar la movilidad de estudiantes entre las instituciones
- Fomentar que los estudiantes se establezcan en las regiones de UNITA
- Crear currículums flexibles y personalizados
- Fomentar el compromiso cívico
- Introducir pedagogías innovadoras
- Lograr un impacto sostenible y un crecimiento azul

[1] Investigador Asociado de la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC)

Resultados Obtenidos

- El proyecto “Diseño y seguimiento de actividades educativas colaborativas online” de la Facultad de Informática de la UNLP ha sido premiado en la convocatoria “Premios a las Mejores Prácticas” enmarcado en el Programa GEMINAE UNITA. El mismo fue presentado como parte del área temática de Enseñanza y aprendizaje y métodos innovadores en el aprendizaje. Forma parte de las investigaciones que se realizan en el Instituto de Investigación en Informática LIDI de la Facultad de Informática de la Universidad Nacional de La Plata, y su aplicación en un seminario de postgrado de esta institución.

  El proyecto se enfoca en aspectos clave del diseño y seguimiento de actividades educativas colaborativas mediadas por tecnologías digitales. Propone estrategias e indicadores clave para poder reflejar (mediante mirroring) el proceso colaborativo, considerando la participación individual y la grupal.
- Se obtuvo un Erasmus KA171 para movilidad y fortalecimiento de los temas de interés de la red. Estas acciones se ejecutarán entre 2025 y 2026. Asimismo se buscará sostener este tipo de presentaciones para potenciar las acciones del programa.

Formación de Recursos Humanos

En vinculación con la Universidad de Zaragoza, que ha sido el nexo con el Programa GEMINAE UNITA se cuenta con una tesis doctoral de co-tutela en temas de interés conjunto. Se está avanzando además en la co-dirección de dos tesis de maestría. Se focalizan acciones de vinculación con el medio y transferencia en función de temas de interés de las universidades que participan en este programa. En particular, en relación a la Universidad de La Plata, la de Zaragoza y el resto de las universidades participantes de Latinoamérica.

Referencias

- Sitio web Programa UNITA GEMINAE. https://univ-unita.eu/Sites/UNITA/en/
- Astudillo, G., Asiain Ansorena, D., Sanz, C., Ponce De León, J., Beltrán Blázquez, J.R. and Baldassarri. S. 2024. Capturing physiological data for detecting boredom in educational contexts. In Proceedings of the XXIII International Conference on Human Computer Interaction (Interacción '23). Association for Computing Machinery, New York, NY, USA, Article 27, 1–5. https://doi.org/10.1145/3612783.3612811 **York, NY, USA, Article 27, 1–5. https://doi.org/10.1145/3612783.3612811**
- Ferreyra, P., Coma-Roselló, T., & Sanz, C. V. (2024). Revisión sistemática sobre videojuegos como método de evaluación sigilosa para analizar rasgos de personalidad. In XIX Congreso Nacional de Tecnología en Educación y Educación en Tecnología-TE&ET 2024 (Morón, 13 y 14 de junio de 2024).

RED DAI: Curaduría y gestión de Datos Abiertos de Investigación: una mirada desde el aporte de la Educación Superior Latinoamericana a la Ciencia Abierta

Ana Casali[1,2], Claudia Deco[1], Paola Bongiovani[3], Regina Motz[4], Elaine Oliveira Lucas[5]

[1] Facultad de Ciencias Exactas, Ingeniería y Agrimensura,
Universidad Nacional de Rosario, Argentina (UNR)
{acasali, deco}@fceia.unr.edu.ar
[2] Centro Internacional Franco Argentino de Ciencias de la Información y de Sistemas (CIFASIS)
[3]Facultad de Humanidades y Artes (UNR)
[4]Facultad de Ingeniería, Universidad de la República, Uruguay (UdelaR)
[5]Universidade do Estado de Santa Catarina, Brasil (UDESC)

RESUMEN

La ciencia abierta ha surgido en la última década como una tendencia poderosa de cambio en la política de investigación y alcanza a todo el ciclo del proceso científico. El concepto de ciencia abierta se basa en concebir el conocimiento que ésta genera, como bien común; sus resultados, gestión y gobierno, tienen efectos sobre el bienestar de la población y su desarrollo. En este marco, poner a disposición o indicar repositorios de datos de investigación y asesorar a sus investigadores sobre dónde, cuándo y cómo depositar sus datos, se ha vuelto un requisito exigido por los organismos de financiación, tanto a nivel regional como internacional. En este sentido, este proyecto asume que cada universidad o centro de investigación necesitará, a corto o mediano plazo, un repositorio para almacenar los volúmenes de datos generados durante el desarrollo de una investigación, capacitar a sus investigadores en el uso y carga de estos repositorios, así como en el desarrollo del Plan de Gestión de Datos (PGD) que cubra los aspectos de almacenamiento, vigencia y privacidad de los datos. La propuesta de esta Red pretende compartir prácticas, experiencias y conocimientos científicos desarrollados por Uruguay, Argentina y Brasil. Esto permitirá fortalecer la colaboración en educación superior en el MERCOSUR para establecer estrategias de curación y gestión de datos abiertos de investigación para el desarrollo de la ciencia abierta en América Latina. Para ello, se impulsa la formación y consolidación de una red académica para la producción de conocimiento sobre el tema, fortaleciendo los grupos de investigación ya existentes y promoviendo el diálogo académico regional a través de la investigación conjunta y la movilidad de investigadores.

Palabras clave: Ciencia Abierta, Datos Abiertos de Investigación, Educación Superior.

CONTEXTO

"Red DAI: Curaduría y gestión de Datos Abiertos de Investigación: una mirada desde el aporte de la educación superior latinoamericana a la Ciencia Abierta" aprobada en la III Convocatoria - Redes Académicas de Investigación NEIES (Núcleo de Estudios e Investigaciones en Educación Superior) del MERCOSUR (2023-2025). Coordinadora por la Dra. Regina. Motz de UdelaR (Universidad de la República, Uruguay) e integrada por investigadores de la UNR y de la UDESC (Universidade do Estado de Santa Catarina (Brasil).

1. INTRODUCCIÓN

La *ciencia abierta* ha surgido en la última década como una tendencia poderosa de cambio en la política de investigación y alcanza a todo el ciclo del proceso científico. El concepto de ciencia abierta se basa en concebir el conocimiento que esta genera, como bien común; sus resultados, gestión y gobierno, tienen efectos sobre el bienestar de la población y su desarrollo (Albornoz et. al, 2019). La ciencia abierta se orienta a que todas las etapas del trabajo científico sean transparentes, ampliamente colaborativas y accesibles. Esto idealmente promueve la reproducibilidad de la ciencia, preserva la integridad de la investigación, evita la duplicación de esfuerzos y promueve la apropiación social de sus resultados. Trabajar de modo abierto puede hacer que la práctica científica sea más efectiva y en gran medida incrementar la variedad y extensión del conocimiento y su acceso. Este enfoque se caracteriza por la apertura no solo de las publicaciones (open access), sino también de la investigación, los datos, metodologías, prácticas y procesos, y la participación de los ciudadanos en un entorno de investigación e innovación responsable. Se trata de poner a disposición los resultados y los procesos de la investigación financiada con fondos públicos para que sean accesibles para la comunidad científica que los produce y para la sociedad en general que los financia, promoviendo la reproducibilidad de la ciencia y la apropiación social de sus resultados. Ante estos desafíos, es necesario desarrollar habilidades relacionadas con la ciencia abierta, así como crear políticas y plataformas que impulsen su integración (O'Carroll et al., 2017; Unesco, 2021).

Según (Tennant et al., 2019), es posible identificar diez componentes del ecosistema abierto: principios abiertos, colaboración abierta, investigación y análisis de datos reproducibles, datos de investigación abiertos, código abierto y software de investigación abierto, acceso abierto a artículos de investigación, evaluación abierta, compromiso público con la ciencia, recursos educativos abiertos y promoción abierta. En particular, la propuesta de esta red de investigación se focaliza en las componentes de análisis de datos reproducibles y repositorios para datos de investigación abiertos, aportando a la construcción de sistemas de gestión que atiendan la curaduría de datos para análisis reproducibles y den asistencia para la gestión de los datos en todo su ciclo de vida.

Para este propósito están surgiendo numerosos repositorios que aceptan una amplia gama de tipos de datos en una variedad de formatos, sin intención de integración o armonización y con pocos requisitos en su descriptores (Wilkinson et al., 2016). Entre algunos servicios e iniciativas de gran interés para facilitar esta perspectiva de ciencia abierta en Europa, se pueden mencionar los siguientes: el Centro de Ciencia Abierta y el Marco de Ciencia Abierta (OSF); Dataverse, una aplicación web de código abierto para compartir, preservar, explorar y analizar datos de investigación; OpenAIRE, una iniciativa para cambiar la comunicación académica hacia la apertura y la transparencia y facilitar formas innovadoras de comunicar y monitorear la investigación; Figshare, servicio web diseñado para académicos, gestión de datos de investigación y difusión de datos de investigación.

En América Latina, no se han formulado aún propuestas globales para la región. Cada país establece su alineación con la ciencia abierta y la materialización del enfoque es desigual. Investigaciones en Latinoamérica respecto a Ciencia Abierta y Datos abiertos pueden verse en (Casali et al, 2022, Rodés, V. & Motz, R., 2022). Estas diferencias se perciben también en el MERCOSUR. La transferencia de las propuestas a las prácticas en las instituciones locales se enfrenta a contradicciones y múltiples factores que inciden en una adopción efectiva. Un avance muy importante lo constituye LA Referencia (Red Federada de Repositorios Institucionales de Publicaciones Científicas), que refleja una voluntad política de ofrecer acceso abierto a la producción científica proveniente de América Latina y que desde 2022, viene realizando foros con el

tema de gestión de datos de investigación (datos abiertos). Sin embargo, las propuestas por ahora no cristalizan en herramientas de asistencia directa a la gestión de los datos de investigación, sino que remiten a buenas prácticas y guías. Incluso las plataformas globales como OSF (Open Science Framework) o RDA (Research Data Alliance) se remiten a ofrecer espacios de almacenamiento persistente para las colecciones de datos con procesos de creación asociados a descripciones en archivos persistentes pero sin controles de calidad sobre la gestión de los datos.

Para promover el desarrollo de Datos Abiertos en la región, en esta Red identificamos cuatro áreas de trabajo: 1) Construcción de una comunidad involucrada con el paradigma de ciencia abierta; 2) Investigación sobre sistemas de información que permitan dar soporte a una curaduría y gestión de datos de investigaciones de calidad; 3) Promoción de materiales educativos, diseño y dictado de cursos, seminarios y talleres que den formación en los conceptos de ciencia abierta; y 4) Desarrollo y promoción de software de código abierto para apoyar la gestión de los datos reproducibles y de asistencia para la gestión de los datos abiertos de investigaciones en todo su ciclo de vida.

Para ello, se pretende impulsar la formación y consolidación de una red académica para la producción de conocimiento sobre el tema, fortalecer los grupos de investigación ya existentes en las tres universidades, promover el diálogo académico regional a través de la investigación conjunta y la movilidad de investigadores, desarrollar herramientas que incorporen técnicas de inteligencia artificial, ayudar a la comunidad académica en la curación y gestión de datos, e impulsar la comunicación científica desde una perspectiva de ciencia abierta latinoamericana.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El objetivo general de esta propuesta es consolidar un enfoque abierto de la ciencia a través de una iniciativa multimodal que permite el acceso a recursos distribuidos para la gestión de los datos de investigaciones científicas, desde la recolección, curaduría, exploración y publicación de datos de investigación, teniendo como caso de estudio investigaciones en ciencias educativas digitales. Desglosamos a continuación este objetivo general en objetivos para cada una de las áreas identificadas:

1. Construcción de una comunidad

La ciencia abierta necesita la participación de múltiples partes interesadas: investigadores, docentes, estudiantes, personal universitario (donde los especialistas en ciencias de la información y bibliotecarios juegan un papel importante), pero más ampliamente, necesita llegar a ser un concepto del cual se apropie toda la sociedad. Esto significa que la divulgación y la participación es un factor decisivo para el éxito. El objetivo es contribuir a una definición compartida y bien entendida de ciencia abierta, promoviendo el diálogo académico en el MERCOSUR a través de la investigación conjunta y la movilidad de investigadores, con foco en la importancia de la Ciencia Abierta en la práctica de la Educación Superior.

2. Investigación

La Red tiene como objetivo desarrollar la participación de sus integrantes en un proyecto de investigación compartido a través de actividades de I+D que aporten conocimiento científico y técnico para la innovación, práctica profesional y académica, trabajando desde el enfoque centrado en el investigador, integrando rigurosamente métodos interpretativos con métodos de inteligencia artificial. Esta investigación se referirá a todos los aspectos de los sistemas de información que atiendan la curaduría y la gestión de los datos de investigación abiertos.

3. **Enseñanza**

En esta área se propone: colaborar en la curaduría, creación, adaptación así como traducción, y difusión de materiales educativos; incluir cursos sobre ciencia abierta dirigidos a diferentes audiencias cubriendo los aspectos de gestión en relación con la metodología de investigación, la importancia de trabajar bajo estándares de ciencia abierta, publicar datos de investigación en formato abierto y utilizar repositorios de datos abiertos de investigaciones; realizar formación de recursos humanos calificados en ciencia abierta a través de la realización de cursos, talleres y mentorías.

4. Software

Se propone desarrollar, promover y poner a disposición una infraestructura abierta compartida para la cooperación, así como herramientas de código abierto que incorporen técnicas de inteligencia artificial y asistan a la comunidad académica en la curaduría y gestión de datos abiertos de investigaciones.

3. RESULTADOS

Las actividades principales que se han realizado en el contexto de la Red DAI son:

Organización de eventos y encuentros:

Encuentros virtuales quincenales de las coordinadoras de la Red y encuentros virtuales con nuevos integrantes de los distintos equipos que se han incorporado al trabajo de la Red.

"Primer Encuentro de la Red DAI" realizado en la UDESC, Florianópolis, Brasil del 11 al 14 de marzo de 2024 en el cual se realizaron las siguientes actividades: encuentro con equipo de Bibliotecólogos UDESC y UFSC, encuentro con equipo de Ciencias de la Educación UDESC, y Jornada "1er. Encontro da Red DAI" realizada en la UDESC el 12-3-2024 para toda la comunidad universitaria

"1er. Encuentro Iberoamericano de Ciencia Abierta y Educación Abierta" como evento asociado de la "XIX Conferencia Latinoamericana de Tecnologías de Aprendizaje - LACLO 2024" realizado en octubre de 2024 en Montevideo (Uruguay). Conjuntamente se realizó el "2ª Encuentro de la Red DAI".

Difusión de los temas centrales de la Red y objetivos de la misma en diferentes eventos:

Presentación "Rede DAI: uma proposta a partir da contribuição da Educação Superior latino-americana para a Ciência Aberta", CONFOA 2023: 14 Conferência Lusófona de Ciencia Abierta, realizada en Natal, Brasil, E. Lucas, setiembre 2023 (Motz et al., 2023).

Presentación de la Red DAI en las IX Jornadas Binacionales de Educación Superior (Brasil-Uruguay), septiembre 2023 en Rivera, en la Mesa "Educación digital abierta y ciencia abierta", R. Motz.

Seminario en el Centro Internacional Franco-Argentino de Cs. de la Información y de Sistemas (CIFASIS: CONICET-UNR): "Ciencia ciudadana: Datos abiertos de Investigación y Didáctica de las Ciencias de la Computación", Ana Casali, septiembre 2023.

Presentación de Red DAI en "1er. Encuentro Iberoamericano de Ciencia Abierta y Educación Abierta", Montevideo, Uruguay, octubre de 2024.

Incorporación de nuevos integrantes y vinculación con otras redes

Se incorporaron a la Red DAI: Cristian Cechinel, Universidade Federal de Santa Catarina (Brasil), Fabiano Couto, Universidade Federal de Rio Grande do Sul; IBICIT (Brasil) y Laura Rezende, Universidade Federal de Goiás; IBICIT (Brasil)

Vinculación de la Red DAI con la Red Española REUNI+D (Red Universitaria de Investigación e Innovación Educativa), organizando en conjunto el "1er. Encuentro

Iberoamericano de Ciencia Abierta y Educación Abierta".

Estancias de investigación

Gonzalo Torterolo (UDELAR) en la UDESC en marzo de 2024, donde se avanzó en su trabajo de su tesis de grado sobre repositorios de datos abiertos de investigaciones.

Elaine Lucas (UDESC) en UDELAR del 13 y 17 de abril 2024, participó con una ponencia en el Coloquio Interdisciplinario para la construcción de la Educación Digital Abierta.

Elaboración de un nuevo proyecto

Se trabaja en la propuesta de un nuevo proyecto sobre Educación Abierta y Ciencia Abierta, junto a un gran consorcio de universidades entre las que se encuentran: Universidad de Santiago de Compostela (USC, España), Universidad Nacional de Rosario (UNR, Argentina), Universidad Nacional de La Plata (UNLP, Argentina), Universidade do Estado de Santa Catarina (UDESC, Brasil), Universidade de Brasília (UnB, Brasil), Laboratorio Nacional de Informática Avanzada (LANIA, A.C., en México), Benemérita Universidad Autónoma de Puebla (BUAP, México), Universidad de la República (Udelar, Uruguay), Universidad Tecnológica (UTEC, Uruguay), Universidade Aberta (UAb, Portugal) y Institute for the Future of Education Europe (IFE, Europe).

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo está integrado por las Dras Ana Casali, Claudia Deco y Paola Bongiovani (UNR), Dra. Elaine Lucas (UDESC) y Dra. Regina Motz (UdelaR). Además, se ha defendido una tesis doctoral vinculada al proyecto (Paola Bongiovani) y se está realizando una Tesis de Grado (Gonzalo Torterolo).

5. BIBLIOGRAFIA

Motz, R.; Casali, A.; Oliveira Lucas, E.; Deco, C. Rede DAI: uma proposta a partir da contribuição da Educação Superior latino-americana para a Ciência Aberta. BiblioCanto, [S. l.], v. 9, n. 2, 2023.

Albornoz, D., Hillyer, B., Posada, A., Okune, A., & Chan, L. (2019). Principles for an inclusive open science: the OCSDNet Manifesto. CONTEXTUALIZING OPENNESS, 23.

Casali, A., Motz, R., & Sprock, A. S. (2022, October). Open Data Management of Research in Education: knowledge and practices in LATAM. In 2022 XVII Latin American Conference on Learning Technologies (LACLO) (pp. 1-7). IEEE.

O'Carroll, C, Kamerlin, C., Brennan, N., Hyllseth, B., Kohl, U., O'Neill, G., & Van Den Berg, R. (2017). Providing researchers with the skills and competencies they need to practise Open Science. Luxembourg: European Commission - Publications Office of the European Union. http://ec.europa.eu/research/openscience/pdf/os_skills_wgreport_final.pdf #view=fit&pagemode=none

Rodés, V. & Motz, R. (2022). Interdisciplinary constructions in Open Education and Science. Informatio [online]. 2022, vol.27, n.1, pp.142-166. Epub 01-Jun-2022. ISSN 2301-1378.

Tennant, J., Becker, B., Bie, T. de, Colomb, J., Goglio, V., Grigorov, I., Hartgerink, C., Hartley, R., Havemann, J., Kramer, B., Madan, C., Masuzzo, P., Matthias, L., Schlatter, M., Steiner, T., & Vos, R. (2019). What Collaboration Means to Us: We are more powerful when we work together as a community to solve problems. Collaborative Librarianship, 11(2).

Unesco. (2021). Proyecto de recomendación sobre la ciencia abierta. https://unesdoc.unesco.org/ark:/48223/pf0000379949_spa

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3(1), 160018.

UNI-UEAR - Alianza Universitaria Argentina Europea para la Transformación Digital
Avances de ejecución

Director del Proyecto: Mario Gimelli - Consejo Interuniversitario Nacional (CIN)
Asesora Académica: Elsa Estevez (UNS)
UBA – UNLP – UNQ – UNS - UNSAM
uniuear@cin.edu.ar

RESUMEN

Se presentan los avances del proyecto internacional que tiene como objetivo contribuir al desarrollo y promoción de los derechos digitales en Argentina, incrementar la cooperación entre Europa y Argentina en temas de Transformación Digital mediante la generación y difusión de conocimiento en el tema, y promover alianzas del sector académico con instituciones del sector público, privado, y de la sociedad civil en temas de transformación digital. Cofinanciado por la Unión Europea, el proyecto se encuentra bajo la coordinación del CIN y participan 5 Universidades Nacionales y 4 Instituciones Europeas. En particular, el artículo describe los objetivos de investigación, actividades a desarrollar y entregables a producir.

Palabras Claves
Transformación Digital

CONTEXTO

El proyecto "*UNI-UEAR - Alianza Universitaria Argentina Europea para la Transformación Digital*" cofinanciado por la Unión Europea, es ejecutado por el Consejo Interuniversitario Nacional (CIN) en colaboración con un consorcio de cinco universidades nacionales: Universidad de Buenos Aires (UBA), Universidad Nacional de La Plata (UNLP), Universidad Nacional de Quilmes (UNQ), Universidad Nacional del Sur (UNS) y Universidad Nacional de San Martín (UNSAM) y cuatro instituciones europeas: European University Institute (Italia), Gdansk University of Technology (Polonia), Technical University of Delft (Países Bajos), y University for Continuous Education Krems (Austria). Tiene como objetivo principal contribuir a la promoción de los derechos digitales en Argentina; y en particular, busca fortalecer la creación y difusión de conocimiento aplicado para la transformación digital; y desarrollar alianzas entre la academia, los sectores público y privado y la sociedad civil para la transformación digital.

El Proyecto se ejecuta bajo el acuerdo NDICI HR INTPA/2022/437-643 -- OPSYS: PC-20191.

1. INTRODUCCION

La transformación digital se define como un proceso donde las tecnologías digitales crean disrupciones que sirven para que las organizaciones den respuestas estratégicas para la creación de valor, mientras, simultáneamente, administran los cambios estructurales y las barreras organizativas que afectan los resultados positivos y negativos del proceso.

Estos cambios requieren de la construcción de nuevas capacidades, tanto humanas, como organizacionales e institucionales. Estos tres aspectos se discuten debajo.

Los recursos humanos de las instituciones deberán trabajar en nuevos entornos, compartiendo sus tareas y responsabilidades con máquinas y robots. Para ello, deberán incorporar nuevos

conocimientos y habilidades. Por una parte, algunos puestos de trabajo, principalmente aquellos relacionados con tareas repetitivas y rutinarias, podrían ser reemplazados por algoritmos, robots y máquinas. Por otro, se requerirán nuevos puestos de trabajo para poder diseñar, desarrollar, desplegar y mantener todos esos productos tecnológicos.

A nivel recursos organizacionales, algunas de las actuales estructuras organizativas no serán ya necesarias y se requerirán otras para gestionar los nuevos escenarios de trabajo. Asimismo, las instituciones contarán con nuevos recursos y herramientas para realizar sus tareas. A fin de ilustrar los cambios en el sector público, se dice que la transformación digital, en particular la automatización de tareas está reinventando los conceptos de gobierno tradicional. Un ejemplo es los burócratas de nivel de calle – representantes de gobierno que directamente interactúan con ciudadanos y hacen decisiones en la calle que afectan a las personas, versus algoritmos de nivel de calle – sistemas automatizados que ya no interactúan con los ciudadanos, sino que en base a los datos disponibles toman decisiones sobre ellos (Alkhatib, A., Bernstein, M., 2019). Ambos, los burócratas y los algoritmos de nivel de calle aplican políticas de gobierno a varias situaciones de decisiones referidas a las personas ciudadanas. Traducen las "políticas definidas" en las "políticas implementadas". Por ejemplo, un oficial de policía realiza una infracción a un conductor, un profesor exime a un estudiante de un prerrequisito para un curso, y un médico prescribe un tratamiento. En todos estos casos, las decisiones impactan directamente en la vida de las personas, pero los burócratas de nivel de calle y los algoritmos de nivel de calle lo hacen de manera diferente y con resultados distintos. Mientras que el burócrata aplica discreción para decidir en los casos usando su experiencia, conocimiento y sensibilidad sobre el contexto; los algoritmos no poseen la capacidad de reflexión humana. Usualmente, toman las decisiones aplicando las reglas que se definieron previamente.

A nivel de capacidades institucionales, los cambios introducidos requieren la revisión y adaptación de los marcos legales y regulatorios vigentes. Se deberán definir los límites de lo que es posible y permitido y lo que viola cuestiones éticas y morales de la sociedad. A modo de ejemplo, decidir qué datos personales se pueden intercambiar e integrar a fin de proveer mejores productos y servicios sin violar la privacidad de las personas; definir las situaciones en las que se permite el uso de un algoritmo que no puede explicar sus decisiones (cajas negras) sin violar los derechos humanos de los individuos; o de qué manera asegurar la equidad de los servicios automatizados cuando los mismos toman decisiones en base a datos que usualmente presentan sesgos para las poblaciones vulnerables (Estevez, E., Janowski, J, Roseth, B,, 2024).

A continuación, explicamos el marco conceptual de la transformación digital definido para este proyecto, el cual se muestra en la Figura 1.

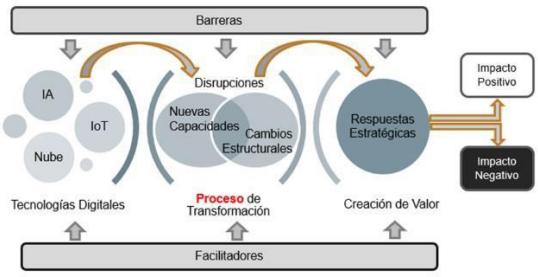


Figura 1

La transformación digital se modela como un proceso que busca crear nuevas capacidades y realizar cambios estructurales (parte central del diagrama), a partir de la utilización de nuevas tecnologías – como la Inteligencia Artificial (IA), computación en la nube, blockchain, Internet de las Cosas, y otras; que se muestran como entradas al proceso (a la izquierda); con el objetivo de producir

respuestas estratégicas, que dependerán específicamente del tipo de institución, para la creación de valor (salida del proceso mostrado a la derecha). Dicha creación de valor dependerá considerablemente de las capacidades de los recursos humanos de la institución, que a su vez, son parte de las nuevas capacidades desarrolladas por el proceso de transformación de la institución, tanto sea pública como privada. A su vez, dichas respuestas pueden tener impactos positivos (como mejorar la calidad de servicios prestados por la organización) como negativos (como violar la protección de datos personales y la privacidad de las personas). Por último, las organizaciones y el contexto en las cuales operan ofrecen facilitadores, que permitan implementarlas de manera más sencilla, y barreras, que obstaculizan su desarrollo.

2. OBJETIVOS DEL PROYECTO

El objetivo general del proyecto es contribuir al desarrollo y promoción de los derechos digitales en Argentina. A partir de este, se derivaron tres objetivos específicos:

- Incrementar la cooperación entre Europa y Argentina en temas de transformación digital.
- Generar conocimiento en temas de transformación digital y difundirlo.
- Desarrollar alianzas del sistema académico con diferentes actores públicos y privados, en proyectos y desarrollos vinculados con la transformación digital.

A fin de cumplir con el segundo objetivo se investigará el estado del arte en temas de transformación digital. En particular, se relevará la oferta académica, esto es cursos de grado y posgrado y programas de educación destinados a crear talento para la transformación digital en Argentina y en países de la Unión Europea. Asimismo, se realizó un estudio sobre las necesidades locales en capacitación para la transformación digital. A tal efecto, se organizaron en Argentina, 30 talleres a lo largo de todo el territorio nacional con participación de diferentes actores de la sociedad para recolectar información sobre: 1 - Proyectos de transformación digital en curso y actores involucrados; 2 - Actores a capacitar y tipo de capacitación necesaria para cada uno de ellos; 3 - barreras y desafíos para iniciativas de transformación digital en el territorio.

A partir de las demandas de capacitación relevadas, actualmente está en desarrollo un conjunto de cursos (en castellano). Estos cursos se dictarán a distancia al menos una vez cada uno por las universidades miembro del consorcio. Los mismos quedarán en el dominio público, a disposición de las universidades argentinas para que sus contenidos puedan sean reutilizados por las mismas de la manera más conveniente.

Asimismo, se apoyará el desarrollo de proyectos que entreguen soluciones concretas para la transformación digital en las diferentes regiones de Argentina. Los mismos deberán responder a necesidades locales identificadas a través de los talleres realizados y serán ejecutados, preferentemente por universidades involucradas en la realización de los talleres.

3. RESULTADOS OBTENIDOS/ ESPERADOS

- Se ha desarrollado un informe sobre el estado del arte en transformación digital, realizando un relevamiento de las carreras y cursos que se dan en la Argentina y en Europa.
- Se han realizado 30 talleres a lo largo del país para detectar las necesidades de formación en temas de TD.
- Se han definido los 15 cursos sobre transformación digital con participación de las universidades del Consorcio:

- Fundamentos de TD
- Marco regulatorio para la TD
- TD en PyMEs
- Tecnologías para la TD
- Planeamiento estratégico para la TD
- Gestión del cambio
- Gobernanza de datos
- Gestión de la innovación
- IA para la TD
- Ciberseguridad
- Liderazgo para la TD
- Humanidades digitales
- Modelos de negocios digitales y economía digital
- Innovación en educación con herramientas digitales
- Comunicación para la TD

- Se están desarrollando los contenidos para dichos cursos (contenidos abiertos)

Asimismo, se esperan obtener los siguientes resultados:

- Dictado de los cursos al menos para 450 actores
- Proyectos de transformación digital ejecutado por universidades nacionales en todo el país
- Documento sobre lecciones aprendidas sobre vinculación
- Agenda para colaboración en transformación digital
- Desarrollo de capacidades para la transformación digital en las universidades argentinas

4. FORMACION DE RECURSOS HUMANOS

Adicionalmente a los resultados científicos y académicos, el proyecto prevé la formación de recursos humanos. Estos esfuerzos incluyen:

- capacitación en temas de transformación digital en diferentes niveles, según los requerimientos de los distintos sectores que participen de los Talleres del proyecto.
- participación de los integrantes del proyecto en el dictado de asignaturas/cursos de grado/postgrado en las universidades donde desempeñan sus funciones, así como en otras de las universidades involucradas.

5. AGRADECIMIENTOS

Proyecto co-financiado por la Unión Europea. Acuerdo: NDICI HR INTPA/2022/437-643 -- OPSYS: PC-20191

6. BIBLIOGRAFIA

- Viale, G., Understanding digital transformation: A review and a research agenda, en The Journal of Strategic Information Systems, vol 28 (2), pp. 118-144, (2019).
- Groover, M. P, Automatización, en Britannica. https://www.britannica.com/technology/automation, (2020
- Bertot, J., Estevez, E., Janowski, T., Universal and contextualized public services: Digital public service innovation framework, Government Information Quarterly, vol 3(2), pp. 211-222, (2016).
- Estevez, And., Janowski, T., Roseth, B., Gobierno automatizado – ¿Cuándo prospera o fracasa la automatización en el gobierno? Banco Interamericano de Desarrollo, en prensa, (2022).
- Sommer, Lutz, Industrial Revolution - Industry 4 . 0 : Are German Manufacturing SMEs the First Victims of This Revolution ?, in *Journal of Industrial Engineering and Management* vol 8 (5), pp. 1512–32, (2015).
- Rauch, Erwin, Dominik T Matt, Christopher A Brown, Andrew R Vickery, and Walter Towner, Transfer of Industry 4 . 0 to Small and Medium Sized Enterprises, in Transdisciplinary Engineering Methods for Social Innovation of Industry 4.0, IOS Press, https://doi.org/10.3233/978-1-61499-898-3-63, (2018).
- Müller, Julian Marius, Oana Buliga, and Kai-ingo Voigt, Fortune Favors the Prepared : How SMEs Approach Business Model Innovations in Industry 4.0, in *Technological Forecasting & Social Change* vol 132, pp.2-17. https://www.sciencedirect.com/science/article/abs/pii/S0040162517312039?via%3Dihub , (2018).

- Matt, Dominik T., Vladimir Modrak, and Helmut Zsifkovits, eds. 2020. *Industry 4.0 for SMEs*, Macmillan, Palgrave, https://link.springer.com/book/10.1007/978-3- 030-25425-4, (2020).
- Velasquez Villagran, N., y Estevez, E., "Methodological Aspects for the Secure Implementation of Industry 4.0 in SMEs based on Cybersecurity" en edición.
- Lipsky, M, Street-Level Bureaucracy: The Dilemmas of the Individual in Public Service. Russell Sage Foundation, (1983).
- Alkhatib, A., & Bernstein, M., Street-Level Algorithms: A Theory at the Gaps Between Policy and Decisions. CHI 2019, May 4–9, 2019, Glasgow, UK, 1–13. https://doi.org/10.1145/3290605.3300760, (2019).
- Estevez, E., Janowski, T., Roseth, B., Automated Government – When Does Automation Thrive or Flounder in Government? Interamerican Development Bank, https://publications.iadb.org/en/when-does-automation-government-thrive-or-flounder (2024).
- Estevez, E., et.al., Ethics in Digital Public Services, UNESCO Chair on Knowledge Societies and Digital Governance, Universidad Nacional del Sur, https://unescochair.cs.uns.edu .ar/en/activities/r esearch/ethics-in-digital-public-services/, (2020)

· SI ·

Seguridad Informática

Análisis del aporte de la Inteligencia Artificial a la mejora de la Seguridad en ambientes de Internet de las Cosas

Jorge Eterovic; Edith García; Luis Torres;

Instituto de Investigación en Ciencia y Tecnología
Vicerrectorado de Investigación y Desarrollo
Universidad del Salvador.
Lavalle 1854 – C1051AAB -Ciudad Autónoma de Buenos Aires - Argentina

{jorge.eterovic; edith.garcia; luisantonio.torres}@usal.edu.ar

RESUMEN

En la actualidad existe una gran preocupación por la seguridad en Internet de las Cosas (IoT) debido al aumento del número de dispositivos conectados a Internet en los ecosistemas IoT y a la creciente cantidad de datos sensibles que se transmiten a través de ellos.
A medida que aumenta su uso se convierten en objetivos atractivos para los ciberdelincuentes y crece la cantidad de formas de ataque, haciendo necesario implementar cada vez más complejos sistemas de seguridad para proteger a estos dispositivos.
La Inteligencia artificial (IA) podría ayudar a detectar posibles amenazas y vulnerabilidades en los dispositivos de IoT, así como en las redes a las que están conectados. Las técnicas de IA, como el aprendizaje automático, el aprendizaje profundo y el procesamiento del lenguaje natural, permitirían a los sistemas IoT analizar grandes cantidades de datos, obtener información y tomar decisiones inteligentes.
Estas técnicas serían de utilidad a los sistemas IoT para adaptarse a entornos dinámicos, detectar anomalías y automatizar operaciones de seguridad contra ciberataques. Al aprovechar la IA, los sistemas IoT podrían mejorar la detección de ciberataques, identificar patrones de comportamiento anormales, prevenir ataques antes de que ocurran, asegurar la comunicación y proteger la privacidad de los datos, siendo en última instancia una herramienta valiosa para crear un ecosistema IoT más sólido, seguro y resistente.
Además, la IA podría ayudar a mejorar la autenticación y la autorización de los dispositivos de IoT, lo que reduciría el riesgo de ataques de suplantación de identidad y de apropiación ilegítima de datos sensibles, entre otros. De este modo, la inteligencia artificial proporcionaría una capa adicional de seguridad para tales equipos y reforzaría la protección contra las ciber-amenazas.
Este proyecto de investigación analizará soluciones basadas en IA para identificar patrones de comportamiento inusuales en dispositivos IoT en base a que los algoritmos de IA pueden analizar adecuadamente los datos de los dispositivos para establecer un comportamiento de referencia y activar alarmas, detectando ciber-amenazas.

Palabras Clave:
Inteligencia Artificial. Internet de las Cosas. Seguridad en IoT.

CONTEXTO

El Vicerrectorado de Investigación y Desarrollo (VRID), perteneciente a la Universidad del Salvador (USAL), dicta las políticas referidas a la investigación, concibiéndola como un servicio a la comunidad y entendiendo que los nuevos conocimientos son la base de los cambios sociales y productivos. Con el impulso de las propias Unidades Académicas se han venido desarrollando acciones conducentes a concretar proyectos de investigación uni/multidisciplinarios, asociándolos a la

docencia de grado y postgrado y vinculando este accionar, para potenciarlo, con otras instituciones académicas del ámbito nacional e internacional.

La Dirección de Investigación, dependiente del VRID, brinda soporte a las distintas Unidades de Investigación y a sus investigadores para el desarrollo de Proyectos y Programas de Investigación, nacionales e internacionales, como así también, apoyo y orientación de recursos para la investigación.

A ella pertenece el Instituto de Investigación en Ciencia y Tecnología (RR 576/12) en el cual se enmarca este proyecto denominado "Análisis del aporte de la inteligencia Artificial a la mejora de la seguridad en ambientes de Internet de las Cosas", con una duración de 2 años (2025-2026) y que ya ha sido evaluado y aprobado para su realización por Disposición N° 228/2024.

1. INTRODUCCIÓN

Internet de las cosas (IoT) es el paradigma en el que cualquier cosa (dispositivos, objetos, sujetos, etc.) puede interconectarse a través de Internet con la capacidad de interactuar, recopilar, procesar y compartir datos de manera inteligente [1].

Hemos entrado en una nueva era de posibilidades en materia de redes debido al extraordinario crecimiento del Internet de las cosas. Los dispositivos de IoT, desde dispositivos inteligentes hasta sensores industriales y vehículos sin conductor, prácticamente se han apoderado de todos los ámbitos de la vida cotidiana de las personas y de las empresas. A pesar de su comodidad y conveniencia, la proliferación de dispositivos IoT plantea importantes problemas de seguridad para los ecosistemas IoT [2].

Con tantos dispositivos IoT ingresando al mercado, es necesario que haya más funciones de seguridad integradas. Además, las capacidades informáticas de los dispositivos IoT están en duda. Se ha demostrado que los dispositivos IoT son vulnerables a las ciber-amenazas, pero también es fundamental conocer la naturaleza de estas amenazas [3]. Ellas son:

- **Gran variedad de dispositivos**. Innumerables fabricantes producen dispositivos IoT en varios tamaños, formas y capacidades. El sistema operativo y los protocolos de comunicación que utilizan son diversos. La falta de uniformidad plantea un gran problema a la seguridad de los dispositivos de IoT.
- **Limitaciones de recursos**. Los fabricantes de IoT inundan el mercado con dispositivos baratos con baja capacidad de almacenamiento, memoria mínima y bajas capacidades de procesamiento. Debido a estas limitaciones, es importante desarrollar soluciones de seguridad innovadoras.
- **Autenticación deficiente**. Los dispositivos IoT son vulnerables al acceso no autorizado debido a contraseñas extremadamente débiles y predeterminadas. La falta de un mecanismo de autenticación fuerte deja a estos dispositivos a merced de los delincuentes informáticos.
- **Rápido crecimiento y escala**. La cantidad de dispositivos IoT está creciendo a un ritmo explosivo, pero sin seguridad. Cuanto mayor sea el número de dispositivos IoT, más débil será la seguridad. Un número cada vez mayor de dispositivos IoT significa una posibilidad cada vez mayor de ciber-ataques exitosos.
- **Dispositivos desactualizados**. Hay una gran cantidad de dispositivos IoT antiguos que son funcionales, pero necesitan una actualización por parte del fabricante. Sin parches de seguridad, estos dispositivos se convierten en una importante amenaza para la seguridad, ya que actualizarlos es costoso.

El rápido crecimiento en la cantidad de dispositivos conectados a Internet ha llevado a la necesidad de buscar soluciones de seguridad.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

La autenticación y la autorización son dos factores claves para asegurar a los dispositivos de la Internet de las Cosas. Se pueden definir de la siguiente manera:

- **Autenticación**: Se refiere a la verificación de la identidad de usuarios o dispositivos.
- **Autorización**: Consiste en la asignación de permisos y acceso a los propietarios, a los usuarios, a los datos sensibles y a la gestión de la red y de los dispositivos.

Existen diferentes tecnologías emergentes como Blockchain [4], Criptografía Ligera [5], Contratos Inteligentes [6], Tokens No Fungibles (NFT) [7] y la Inteligencia Artificial (IA) [8] que podrían asegurar los procesos de Autenticación y Autorización. En particular, la IA podría ser utilizada para mejorar:

- **Los sistemas de autenticación**. La autenticación multifactorial, por ejemplo, requiere que los usuarios proporcionen varios factores de identificación para acceder a los dispositivos. La IA podría ayudar a detectar y prevenir intentos de suplantación de identidad, mejorando la seguridad de los dispositivos IoT.
- **La autorización al monitorear el tráfico de red**. Para analizar el comportamiento de los usuarios y los dispositivos. La IA puede detectar patrones de comportamiento anormales y alertar a los administradores de red de posibles amenazas o intentos de acceso no autorizado. Esto podría ayudar a prevenir los ataques y a optimizar la seguridad de la red.
- **La gestión de identidades y accesos** (Identity and Access Management - IAM) **en los dispositivos IoT**. La IAM trabaja con la gestión de las identidades de los usuarios y dispositivos y con la asignación de permisos y acceso a los recursos. La IA podría ayudar a mejorar la IAM al analizar grandes cantidades de datos y detectar esquemas de comportamiento anormales. Esto ayudaría a evitar los ataques y, por ende, a hacer más seguros los dispositivos de IoT.

Internet de las cosas (IoT) se caracteriza por la gran cantidad de datos que se generan y manipulan a través de los dispositivos conectados. Sin embargo, la privacidad de los mismos es, actualmente, una preocupación crítica para los propietarios y los usuarios de estos dispositivos.

3. RESULTADOS OBTENIDOS/ESPERADOS.

La Inteligencia Artificial (IA) ofrece en este sentido diferentes técnicas [9]; pudiendo ser utilizada para:

- **Enmascarar los datos sensibles**. Esto se logra al agregar ruido a los datos de tal manera que el rumor oculte cualquier información sensible, pero no afecte significativamente su calidad. De esa manera, los datos se pueden usar en aplicaciones de análisis sin comprometer la privacidad del usuario.
- **Detectar y prevenir intrusiones a la privacidad de los datos**. La IA puede monitorear el tráfico de red y analizar los patrones de acceso a la información. Si detecta un comportamiento sospechoso, está en capacidad de alertar a los administradores de red para que actúen rápidamente y eviten la posible exposición de datos sensibles.
- **Mejorar la seguridad de los dispositivos IoT**. Si los dispositivos IoT se ven comprometidos, pueden ser utilizados para recopilar y transmitir información sensible sin que el usuario lo sepa. La IA puede mejorar la seguridad, detectando estas actividades maliciosas y alertando a los usuarios y a los administradores de red de una posible violación de la privacidad de los datos.

Además de los aspectos de seguridad enumerados, la IA puede mejorar la seguridad de la cadena de suministro. Los dispositivos IoT son fabricados por una variedad de proveedores y, por lo tanto, pueden sufrir de vulnerabilidades en el proceso de distribución. La IA puede ser utilizada para monitorear y analizar sus etapas, detectando posibles amenazas.

Finalmente, la IA puede ayudar a prevenir ataques de Denegación de Servicio Distribuido

(DDoS). Los ataques relacionados con la DDoS son una preocupación común en la IoT. La inteligencia artificial puede ser usada para detectarlos y prevenirlos, mejorando la disponibilidad y el rendimiento de tales dispositivos.

Resumiendo, la seguridad en el Internet de las cosas (IoT) es una preocupación creciente debido al aumento de dispositivos conectados y datos sensibles transmitidos. Respecto a este tema, la Inteligencia Artificial (IA) puede contribuir a través de la detección de amenazas en tiempo real, la mejora de la autenticación y autorización, la protección de la privacidad de los datos, la prevención de los ataques DDoS, etc.

Las soluciones de inteligencia artificial suelen ser autosuficientes para aprender de forma independiente, reconocer comportamientos y distinguir entre comportamientos justificados e injustificados. Basándose en el aprendizaje mediante datos de comportamiento, la IA puede determinar las amenazas a la seguridad.

Este proyecto de investigación analizará soluciones basadas en IA para identificar patrones de comportamiento inusuales en dispositivos IoT en base a que los algoritmos de IA pueden analizar adecuadamente los datos de los dispositivos para establecer un comportamiento de referencia y activar alarmas, detectando ciber-amenazas.

4. FORMACIÓN DE RECURSOS HUMANOS.

El equipo de investigadores pertenece al cuerpo docente de Tecnologías Aplicadas de la Facultad de Ingeniería, específicamente al área de Seguridad Informática, de la Universidad del Salvador.

A este proyecto, se incorporaron dos docentes investigadores con amplia trayectoria académica, un docente investigador con muchos años de desempeño en la industria de TI y tres alumnos de la carrerea de Ingeniería en Informática.

Esto redundará en un aumento del activo académico e investigativo representado por su cuerpo de docentes investigadores y alumnos, como así también conformará las bases para la investigación a futuro desarrollando esta línea de trabajo en la Facultad de Ingeniería.

5. BIBLIOGRAFÍA.

[1] Silviu Ciuta; Securing the Internet of Things (IoT): Cybersecurity of Connected Devices; Ed.: Independently published. 2023.

[2] Sunil Cheruvu, Anil Kumar, Ned Smith, David M. Wheeler; Demystifying Internet of Things Security: Successful IoT Device/Edge and Platform Security Deployment. 1st Edition. Ed.: Apress Open. 2020..

[3] Michael Fagan, Katerina N. Megas, Jeffrey Marron, Kevin G. Brady Jr, Barbara B. Cuthill, Rebecca Herold, David Lemire, Brad Hoehn; NIST SP 800-213A: IoT Device Cybersecurity Guidance for the Federal Government: IoT Device Cybersecurity Requirement Catalog; Ed.: NIST (National Institute of Standards and Technology). 2022.

[4] Jorge Eterovic, Marcelo Cipriano, Edith García, Luis Torres; Marco de referencia para aplicaciones seguras de Blockchain; 8 º Congreso Nacional de Ingeniería Informática Sistemas de Información CONAIISI 2020; Libro de Actas CoNaIISI 2020; ISBN 978-950-42-0202-8. Argentina. 2020.

[5] Marcelo Cipriano, Jorge Eterovic, Edith García, Luis Torres; Criptografía Liviana para Internet de las Cosas e Internet de las Cosas Industrial; XXIII Workshop de Investigadores en Ciencias de la Computación; Libro de Actas WICC 2021; ISBN 978-987-24611-3-3; Argentina. 2021.

[6] Jorge Eterovic, Marcelo Cipriano, Edith García, Luis Torres; Aplicación de los Contratos Inteligentes en Internet de las Cosas; XXIII Workshop de Investigadores en Ciencias de la Computación; Libro de Actas WICC 2021; ISBN 978-987-24611-3-3; Argentina. 2021.

[7] Eterovic, J; Cipriano, M; García, E; Torres, L. Propuesta de uso de tokens inteligentes no

fungibles (NFTs) para representar dispositivos de Internet de las Cosas. 11° CoNaIISI 2023 - Congreso Nacional de Ingeniería Informática / Sistemas de Información. Actas del CoNaIISI. 2023. ISBN 978-987-4417-73-2.

[8] Mazhar, T.; Talpur, D.B.; Shloul, T.A.; Ghadi, Y.Y.; Haq, I.; Ullah, I.; Ouahada, K.; Hamam, H. Analysis of IoT Security Challenges and Its Solutions Using Artificial Intelligence. Brain Sci. 2023, 13, 683. https://doi.org/10.3390/10.3390/brainsci13040683.

[9] Ismail Chahid, Mohammed Benabdellah. Artificial Intelligence approaches for IoT Security: State of Art. International Journal of Advanced Engineering. Research and Science (IJAERS). Peer-Reviewed Journal. ISSN: 2349-6495(P) | 2456-1908(O). Vol-9, Issue-6; Jun 2022.

Auditoría en Amazon Web Services

Alejandro Vazquez, Bruno Roberti, Germán Amico
Grupo AuSegTIC (*Grupo de Auditoria y Seguridad de TICs*)
Departamento de ISI, Facultad Regional Mendoza, Universidad Tecnológica Nacional
(avazquez,broberti)@frm.utn.edu.ar

Resumen
En este artículo se busca determinar el alcance de las herramientas de auditoría que brindan los servicios en la nube del tipo "Infrastructure as a Service (IaaS)" o "Plataform As A Service (PaaS)".

Teniendo en cuenta que la tendencia mundial actual es el traslado de los sistemas hacia servicios en la nube, se observa que existen muy pocos estándares procedimentales públicos para el desarrollo de auditorías bajo este tipo de servicios.

Como objetivo de la investigación, considerando lo mencionado anteriormente, se busca analizar de forma teórica y práctica las capacidades de auditoría que brinda la plataforma Amazon Web Services a través de herramienta "AWS Audit Manager" de la misma, identificando los recursos que brinda para poder realizar una auditoría de seguridad de la información de acuerdo al estándar ISO 27002.

También se busca generar un programa de verificación que permita adaptar de manera rápida los estándares de la industria a los servicios de IaaS.

Palabras Claves
Auditoría, Servicios en la nube, Herramientas de Auditoría, Audit Manager, Controles ISO 27002, Programa de verificaciones.

Contexto
Este trabajo de Investigación está inserto en la línea "Auditoría TICs y COBIT" del Grupo de Auditoria y Seguridad de TICs, asociado al Departamento de Ingeniería en Sistemas de Información, de la Facultad Regional Mendoza, Universidad Tecnológica Nacional. Los objetivos de esta línea son obtener conocimiento de normas internacionales, nacionales y estándares de Auditoría y Seguridad de TIC; capacitación y preparación de especialistas en "Auditoría y Seguridad en TIC"; investigar métodos y procedimientos de Auditoría y preparar programas de Auditoría específicos.

Introducción
En el modelo de infraestructura tecnológica on-premise, se tienen recursos como servidores, con su procesador (CPU), memoria (RAM), espacio de almacenamiento, y elementos de red en una misma ubicación; la cual plantea desafíos para el crecimiento más allá de un punto determinado. Utilizar infraestructura física y centralizada, gestionada por uno mismo, tiene un mayor costo de entrada y a largo plazo puede traer problemas con elementos como el suministro de energía, la escalabilidad, y el mantenimiento.

Esto fue una gran barrera para empresas durante mucho tiempo hasta que las tecnologías necesarias estuvieron maduras para permitir que el modelo llamado de computación en la nube pueda ser ofrecido en forma amplia. Este modelo se puede resumir como entrega bajo demanda de poder de cómputo, almacenamiento de datos, aplicaciones para administrarlos y otros recursos de TI. Es decir que permite ofrecer servicios de computación a través de internet como servicio.

Dentro de la computación en la nube existen tres opciones de servicio, como se observa en la Figura 1, clasificados según los servicios ofrecidos y modelos implementados.



Figura 1. Clasificación de servicios.

Dentro de los servicios en la nube se pueden encontrar muchos prestadores, pero uno de los más destacados es AWS (Amazon Web Services)[1]. Se puede observar una comparativa en la Figura 2. Ti on-premise vs AWS.



Figura 2. TI on-premise vs AWS.

Amazon Web Services (AWS) es una plataforma que ofrece un amplio conjunto de productos globales basados en la nube. Un entorno AWS es flexible por lo que puede volver a configurarse y actualizarse bajo demanda, escalarse de forma ascendente automáticamente para cumplir con los patrones de uso y optimizar gastos, o bien, cerrarse de forma temporal o permanente. La facturación de los servicios de AWS se convierte en un gasto operativo en lugar de un gasto capital.

Los servicios de AWS están diseñados para trabajar en conjunto a fin de permitir prácticamente cualquier tipo de aplicación o carga de trabajo. Hay que ver estos servicios como bloques de creación, que pueden juntarse rápidamente para crear soluciones sofisticadas y escalables y, luego, ajustarlas a medida que cambien las necesidades según la situación. Se puede ver algunos de los servicios que se ofrecen en la Figura 3. Servicios AWS.



Figura 3. Servicios AWS.

Como se puede ver, la computación en la nube es una de las innovaciones tecnológicas más beneficiosas de los últimos tiempos. Sin embargo, eso no significa que no haya algunos riesgos dentro de estos sistemas a los que deba hacerse frente, la seguridad es uno de los principales temas a tener en cuenta en los sistemas que hacen uso de la computación en la nube. Basarse totalmente en Internet incrementa la vulnerabilidad a ataques de hackers. Pero la evidencia habla por sí sola, todos los sistemas de TI modernos de hoy en día están conectados a Internet.

Aquí es donde entra la Auditoría en general y específicamente en este caso para los servicios en la nube, que para este trabajo la consideramos como el examen de los controles de seguridad asociados a los servicios y proveedores de nube para satisfacer los objetivos de control definidos en estándares internacionales, brindando certeza al negocio de que se mitigan los riesgos y se maximizan los beneficios de la nube y verificar que se apegan a los mismos.

Existen varios beneficios al realizar una auditoría en la nube:

- Cumplir con las políticas y los marcos de gobierno existentes relacionados con los ambientes de nube.

- Definir la responsabilidad compartida derivada de la interacción de los agentes que intervienen en un esquema de servicio de nube
- Implementar un marco de controles de seguridad específico para la nube para garantizar la seguridad dentro de una organización.
- Integrar los entornos de la nube a la auditoría de TI on-premise.
- Garantizar la seguridad en la nube antes y durante la prestación del servicio utilizando métodos y técnicas de evaluación aceptados.
- Medición de la eficacia de los controles a través de métricas y supervisión continua.

Auditoría en AWS

En base al relevamiento realizado sobre herramientas de auditoría que existen en este momento para servicios en la nube de AWS, se destacan tres en particular, las cuales aportan diferentes enfoques:

"Amazon CloudWatch"[5] es un servicio de monitoreo y observabilidad creado por los ingenieros DevOps, los desarrolladores, los ingenieros de fiabilidad de sitio (SRE) y los administradores de TI. CloudWatch puede monitorear recursos AWS en tiempo real, recopilar y hacer un seguimiento de las métricas, que son las variables referidas a sus recursos y aplicaciones que se pueden medir.

"AWS CloudTrail"[6] que permite realizar auditorías de gobernanza, de conformidad, operativas y de riesgo en la cuenta de AWS. Además se puede registrar, monitorear de manera continua y retener la actividad de la cuenta relacionada con acciones en toda su infraestructura de AWS. Sumado a lo mencionado anteriormente, CloudTrail proporciona el historial de los eventos de actividad realizada en la cuenta de AWS, incluyendo las acciones efectuadas a través de la consola de administración, los SDK de la misma, las herramientas de comandos y otros servicios.

"Amazon Audit Manager"[4] que ayuda a realizar auditorías continuas en el uso de los servicios utilizados de AWS para simplificar cómo administrar el riesgo y el cumplimiento de las regulaciones y estándares de0 la industria. Automatiza la recopilación de pruebas para facilitar la evaluación de las políticas, procedimiento y actividades funcionan de manera eficaz.

Se analizaron las distintas alternativas de herramientas para la realización de la actividad y de dicho análisis, se llegó a la conclusión que la mejor alternativa era utilizar Amazon Audit Manager ya que se presenta como la más .

Metodología

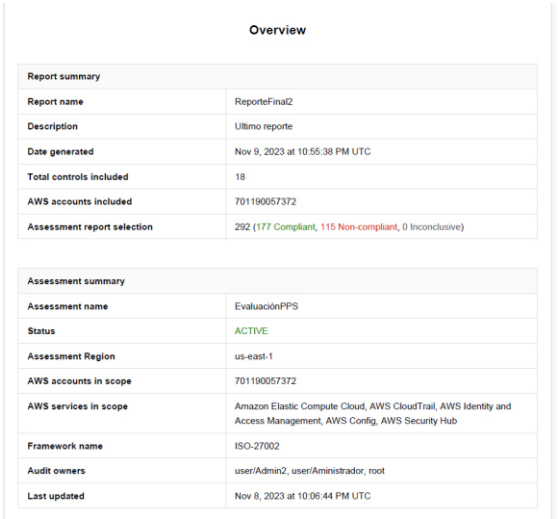
El trabajo se realizó utilizando Amazon Audit Manager, sobre un servicio EC2[2] (IaaS) de AWS, y así se pudo verificar la implementación de controles del estándar de seguridad ISO 27002[3] y su cumplimiento.

Audit Manager trabaja creando evaluaciones con diferentes frameworks de estándares internacionales con controles que ya vienen cargados en la herramienta, o sino también se pueden crear los mismos de forma personalizada, combinando controles según la necesidad que se presente. En este caso los controles de la ISO 27002 están incluidos, pudiendo de esta manera configurar una evaluación en base a los mismos en la propia cuenta de AWS y en específico en la instancia EC2. Una vez creada la evaluación empieza a recolectar evidencias, las cuales serán accesibles luego de 24 horas

Una vez que la evaluación obtuvo suficientes evidencias se realizaron reportes a través del apartado "Evidence Finder" en el cual se pudo filtrar las evidencias obtenidas de la evaluación existente a través de filtros de rango de días, tipo de evidencia y también pudiendo agrupar por diferentes tipos como: ID de cuenta, control, dominio del control, tipo de fuente de datos, nombre de evento,

servicio, tipo de recurso y categoría de servicio.

Los reportes, se presentan en formato pdf separando las evidencias de los controles en carpetas que corresponden a los dominios a los que pertenecen para el caso de la evaluación que se realizó en base a la ISO 27002. También hay un pdf de presentación que enlaza todas las evidencias recolectadas llamado “AssessmentReportSummary”.



Overview

Report summary	
Report name	ReporteFinal2
Description	Ultimo reporte
Date generated	Nov 9, 2023 at 10:55:38 PM UTC
Total controls included	18
AWS accounts included	701190057372
Assessment report selection	292 (177 Compliant, 115 Non-compliant, 0 Inconclusive)

Assessment summary	
Assessment name	EvaluaciónPPS
Status	ACTIVE
Assessment Region	us-east-1
AWS accounts in scope	701190057372
AWS services in scope	Amazon Elastic Compute Cloud, AWS CloudTrail, AWS Identity and Access Management, AWS Config, AWS Security Hub
Framework name	ISO-27002
Audit owners	user/Admin2, user/Aministrador, root
Last updated	Nov 8, 2023 at 10:06:44 PM UTC

Figura 6. Resumen del reporte.

También se puede ver una tabla de contenidos (Figura 7) a modo de índice que lleva a las evidencias de los diferentes controles.



Overview / Table of contents

Table of contents

Overview 2

A.9.1 Business requirements of access control (2 Controls)

A.9.1.1 Access control policy 4

A.9.1.2 Access to networks and network services 6

A.9.2 User access management (3 Controls)

A.9.2.1 User registration and de-registration 10

A.9.2.2 User access provisioning 13

A.9.2.3 Management of privileged access rights 16

A.9.4 System and application access control (3 Controls)

A.9.4.1 Information access restriction 19

A.9.4.4 Use of privileged utility programs 22

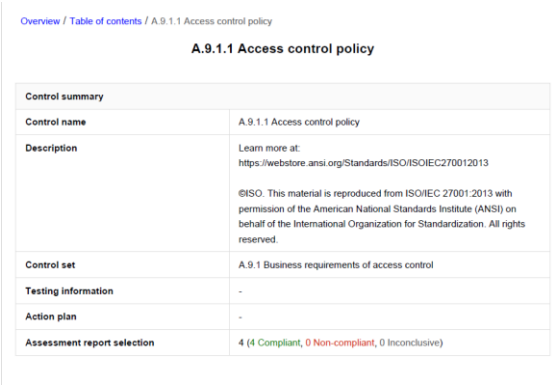
A.9.4.5 Access control to program source code 26

A.12.1 Operational procedures and responsibilities (2 Controls)

A.12.1.2 Change management 29

A.12.1.3 Capacity management 31

Figura 7. Tabla de contenidos.



Overview / Table of contents / A.9.1.1 Access control policy

A.9.1.1 Access control policy

Control summary	
Control name	A.9.1.1 Access control policy
Description	Learn more at: https://webstore.ansi.org/Standards/ISO/ISOIEC270012013 ©ISO. This material is reproduced from ISO/IEC 27001:2013 with permission of the American National Standards Institute (ANSI) on behalf of the International Organization for Standardization. All rights reserved.
Control set	A.9.1 Business requirements of access control
Testing information	-
Action plan	-
Assessment report selection	4 (4 Compliant, 0 Non-compliant, 0 Inconclusive)

Líneas de Investigación

El trabajo de investigación “Auditoría en Amazon Web Services” surge de la necesidad de expandir el conocimiento de los servicios en la nube y propiciar el uso de herramientas automatizadas para realizar auditorías en entornos de computación en la nube que sigan estándares internacionales, como por ejemplo ISO 27002.

También fue objetivo de este proyecto aumentar la documentación pública y accesible sobre estándares procedimentales para el desarrollo de auditorías bajo estos servicios teniendo en cuenta que la tendencia a nivel mundial es el traslado de los sistemas hacia servicios en la nube.

Conclusiones

El resultado del presente trabajo alcanzó los objetivos planteados, ya que la experiencia con Amazon Audit Manager, resultó exitosa. Se pudo implementar la herramienta, la misma permitió verificar en forma directa una amplia variedad de controles propuestos por la normativa ISO 27002; adaptados a un entorno de IaaS. Además el producto presentó evidencia apta para su utilización en un procedimiento estándar de auditoría lo cual permitió la creación de un programa de verificaciones que se adecue a este entorno.

Por lo expuesto Amazon Audit Manager es una herramienta versátil para la realización de las auditorías en el entorno de AWS, nos permite llevar adelante una vasta customización y presenta varios beneficios:

Comenzar rápidamente
Permite seleccionar de entre las plantillas ya desarrolladas en una galería dispuesta por Amazon que soporta un amplio rango de opciones de cumplimientos de estándares y regulaciones. Una vez seleccionada, comienza a automatizar la recolección de pruebas para auditar el uso de los servicios de AWS.
Soporta cumplimientos estándar y regulaciones comunes
Se provee un control del mapeo prediseñado para el cumplimiento de estándares y regulaciones que incluyen CIS Fonundation Benchmark, PCI DSS, GDPR, HIPAA, HITRUST, SOC2, GxP y las mejores prácticas operacionales de AWS.
Framework customizable
Permite crear nuestro propio conjunto de herramientas con los estándares que se desea cumplir y sus respectivos controles. Esta característica permite lograr los requerimientos únicos que son necesarios para auditorías internas.
Soporte para colaboración entre equipos
Se delega el control a expertos de la materia que pueden analizar las pruebas, añadir comentarios y actualizar el estado en cada control.
Crear reportes para auditores externos
Se generan informes de evaluación para resumir las pruebas relevantes recopiladas para nuestra auditoría y se crean carpetas que contienen dichas pruebas de manera detallada.
Asegura la integridad de las pruebas
Se almacenan las pruebas en un lugar seguro donde se mantienen inalterados.

Formación de RRHH

El presente trabajo busca la formación de Recursos Humanos para el área de auditoría de entornos cloud computing, para el desarrollo de actividades de investigación y servicios.
Se inicio dentro del grupo con la llegada de becarios del proyecto “Becas Belgrano” en una etapa inicial y se plasmó en el trabajo de práctica profesional supervisada titulado “Auditoría en Amazon Web Services”

Referencias

[1] AWS: https://aws.amazon.com/es/

[2] EC2: https://aws.amazon.com/es/ec2/

[3] ISO 27002: https://www.iso27000.es/iso27002.html

[4] AWS Audit Manager: https://docs.aws.amazon.com/es_es/audit-manager/latest/userguide/what-is.html

[5] CloudWatch: https://docs.aws.amazon.com/es_es/AmazonCloudWatch/latest/monitoring/WhatIsCloudWatch.html

[6] CloudTrail: https://docs.aws.amazon.com/es_es/awscloudtrail/latest/userguide/cloudtrail-user-guide.html

FRAMEWORK DE AUTOMATIZACIÓN DE INTERCEPCIÓN DE TRÁFICO EN APLICACIONES MÓVILES BASADO EN FRIDA

Fabian A. Gibellini, Leonardo R. Ciceri, Ileana M. Barrionuevo, Juliana M. Notreni, German N. Parisi, Ninfa M. Zea Cardenas, Analia L. Ruhl, Marcelo J. Auquer, Federico Bertola, Sergio R. Quinteros, Ignacio J. Sanchez Balzaretti

GISSIC - Grupo de Investigación en Seguridad de Sistemas de Información y Ciberseguridad. / Laboratorio de Sistemas / Dpto. de Ingeniería en Sistemas / Universidad Tecnológica Nacional / Facultad Regional Córdoba
Cruz Roja S/N, CP: 5016

fabiangibellini@gmail.com, leonardorciceri@gmail.com, ilebarrionuevo@gmail.com, julinotreni@gmail.com, germannparisi@gmail.com, milyzc@gmail.com, analialorenaruhl@gmail.com, marcelo.auquer@gmail.com, fedebertola@gmail.com, ser.quinteros@gmail.com, ignaciojsb@gmail.com

RESUMEN

En este trabajo se presenta una herramienta unificada que reúne diferentes técnicas manuales y automatizadas para facilitar la intercepción de tráfico en aplicaciones móviles. Si bien ya existen soluciones que brindan mecanismos para lograr la captura de tráfico web, con la herramienta propuesta se logra cubrir un abanico de posibilidades de acuerdo a las situaciones que surjan al momento de inspeccionar una aplicación móvil, ya que no todas las herramientas existentes presentan cobertura completa.

Puntualmente, se trabaja con la integración de las herramientas Frida y objection, que permiten la manipulación de objetos y métodos en tiempo de ejecución de la aplicación.

Palabras clave: Seguridad, Test de penetración, Intercepción, Burp Suite, Android.

CONTEXTO

El análisis dinámico de aplicaciones móviles requiere de destinar esfuerzo a la inspección de tráfico web entre aplicación y servidor para encontrar vulnerabilidades y comprender el funcionamiento de los flujos de la aplicación. Para realizar este proceso, se debe tener en cuenta que las aplicaciones están desarrolladas en distintos lenguajes de programación, con distintos frameworks, para distintas plataformas, y además poseen ciertas protecciones para evitar que un intermediario intercepte esa comunicación entre cliente y servidor. Es en este punto donde los pentesters se encuentran con algunas dificultades que deben sortear manualmente y con algunas herramientas automatizadas, pero éstas no cubren todas las expectativas ya que comúnmente requieren de pasos adicionales y, algunas veces, extensos, para lograr que la intercepción sea exitosa. Es por ello que el proyecto pretende reunir tanto técnicas manuales como automatizadas que sirvan como abanico de posibilidades para el pentester en una sola herramienta.

1. INTRODUCCIÓN

Entre los distintos ataques a las aplicaciones móviles, el proyecto se ha enfocado en uno en particular: es el ataque de tipo man-in-the-middle, que explota el hecho de que el servidor HTTPS envía un certificado con su clave pública al navegador web. Si este certificado no es confiable, toda la comunicación es vulnerable (Callegati et ál, 2009) ya que reemplaza el certificado original que autentica al servidor HTTPS servidor con un certificado modificado. El ataque tiene éxito si el usuario se niega a verificar el certificado cuando el navegador envía una notificación de advertencia. Todo esto corresponde al tipo de análisis activo, en donde el investigador intercepta activamente toda la comunicación de red, pudiendo analizar, interactuar, modificar los datos en ese preciso instante. Aquí se utiliza generalmente un proxy web y todas las llamadas hechas y recibidas por

la aplicación y el servidor pasan a través del mismo (Gupta, 2023).

Previo a lograr la captura de tráfico, es preciso manipular, además, los métodos de SSL pinning que posee la aplicación, para lo cual utilizaremos *Frida* y *objection*, este último en particular para cuando se trabaja con dispositivos no *root*.

2. LÍNEAS DE INVESTIGACIÓN y DESARROLLO

Luego de haber analizado las técnicas y herramientas más apropiadas a incorporar al proyecto, aquellas consideradas primordiales para el avance del mismo han sido:

- BurpSuite: requiere de la descarga del certificado desde su propia interfaz, la modificación de la extensión de dicho certificado, y la posterior instalación manual en el dispositivo (PortSwigger, s.f.).
- Frida: implica la instalación manual y ejecución de un agente en dispositivo rooteado, y luego correr el script para saltar la protección de SSL pinning. Esta herramienta posee cobertura respecto al agente servidor, en dispositivos rooteados únicamente. Para dispositivos no rooteados, posee cobertura únicamente a través de la inyección de un gadget en la aplicación objetivo. Funciona tanto en iOS como en Android (Frida, s.f.).
- Objection: requiere modificar la aplicación, generar otra a la que se le inyecta un gadget de Frida, e instalarla en el dispositivo. Funciona con dispositivos rooteados o no rooteados. Posee cobertura tanto en Android como iOS (SensePost, s.f.).

Se decidió programar el script en lenguaje Python por su versatilidad y compatibilidad con distintos entornos de sistema operativo.

3. RESULTADOS Y OBJETIVOS

La ejecución de estas herramientas presentó algunas fallas en Windows, sobre todo al trabajar con configuraciones específicas relacionadas al subsistema de Linux (WSL), pero se ha logrado correr el programa con éxito teniendo en cuenta que se deben modificar configuraciones de red de WSL para adaptarlo a la red local de Windows, y de esta manera, lograr la identificación del dispositivo móvil que se ha conectado.

Tanto Frida como objection son compatibles con dispositivos móviles reales y con emuladores.

A lo largo de las pruebas realizadas, fue notable la mejora de utilizar un solo script para comenzar con el análisis de tráfico, en lugar de tener que determinar manualmente qué tipo de dispositivo está conectado al equipo (PC), qué arquitectura posee y si está rooteado o no, siendo estos chequeos, dependientes de herramientas separadas, de las cuales se toma la salida para luego continuar con otras herramientas a su vez.

Otro punto de mejora logrado, fue la automatización de la instalación del certificado del proxy a utilizar, que si bien en las pruebas se decidió integrar el correspondiente a BurpSuite, se puede reemplazar por cualquier certificado de otra herramienta de proxy web. Este proceso suele consumir esfuerzo manual, que si bien no es excesivo, sí representa un 15% del flujo de trabajo que se desea automatizar en gran medida.

Además de las herramientas mencionadas, se integraron otras que conforman la base para la interacción con el dispositivo móvil (adb, apktool), y también scripts para deshabilitar los métodos de SSL pinning, configuraciones de red puntuales, y mecanismos para editar archivos de configuración de la aplicación en Android, quedando todos estos procesos ya automatizados.

Se informa además, que el proceso de instalación de todos los paquetes requeridos también está automatizado, por lo que el pentester no debe preocuparse por instalar herramientas por separado, ya que el proyecto ha cubierto esta problemática.

Otro objetivo logrado corresponde a la reducción de problemas de compatibilidad de la herramienta desarrollada, dado el uso de entornos virtuales de Python. Se sugiere trabajar de esta manera, para evitar posibles conflictos

respecto a las dependencias que requiere el proyecto para ser ejecutado.

Como objetivos futuros, el equipo se plantea adaptar el script para comunicación con dispositivos iOS, dado que Frida ya posee dicha compatibilidad, sólo falta incorporar la fase de reconocimiento de dispositivos de este tipo al proyecto.

La herramienta ha sido de gran ayuda durante los penetration tests realizados por los integrantes del grupo GISSIC a determinadas aplicaciones móviles, logrando un ahorro del 70% del tiempo en incurrido en el proceso de instalación y configuración tanto de las herramientas apropiadas, como de la adecuación de las aplicaciones para funcionar con dichos programas.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo está conformado por docentes e investigadores de la Universidad Tecnológica Nacional - Facultad Regional Córdoba (UTN - FRC), integrantes del Laboratorio de Sistemas y alumnos becarios de la carrera de Ingeniería en Sistemas de Información. Los estudiantes tendrán la oportunidad de iniciar su formación en investigación científica y tecnológica, profundizando sus conocimientos en temas como la seguridad de la información y el desarrollo de software.
Los docentes que participan en este proyecto están relacionados con asignaturas del plan de estudios, tales como Seguridad en los Sistemas de Información, Seguridad en el Desarrollo de Software, Sistemas Operativos, Redes de Información, Algoritmos y Estructuras de Datos, y Desarrollo de Software. Además, se brindará a los estudiantes la posibilidad de realizar la Práctica Supervisada.
Además, se prevé organizar charlas, talleres y conferencias tanto a nivel provincial como nacional, con el objetivo de compartir los resultados del proyecto en los ámbitos académico y profesional. Congresos como el WICC, WEEF, JAIIO, JEIN y CONAIISI serán lugares para presentar avances y resultados finales, fomentando la interacción entre académicos, investigadores y profesionales.
Al finalizar el proyecto, los conocimientos y desarrollos generados serán transferidos al Laboratorio de Sistemas del Departamento de Ingeniería en Sistemas de Información, para continuar esta línea de trabajo y fortalecer las capacidades locales en seguridad de la información. Este enfoque permitirá que los egresados de la UTN - FRC no solo adquieran habilidades técnicas avanzadas, sino también una formación orientada a la resolución de problemas reales, impactando positivamente en la industria y en la sociedad.

5. BIBLIOGRAFÍA

Callegati, F., Cerroni, W., & Ramilli, M. (2009). Man-in-the-middle attack to the HTTPS protocol. IEEE Security & Privacy, 7(1), 78-81. https://doi.org/10.1109/MSP.2009.12

Frida. (s.f.). FRIDA - Dynamic instrumentation toolkit for developers, reverse-engineers, and security researchers. Recuperado el 19 de abril de 2024, de https://frida.re/

Gupta, A. (2023). Learning pentesting for Android devices. Packt Publishing.

PortSwigger. (s.f.). Burp Suite Community Edition. Recuperado el 19 de abril de 2024, de https://portswigger.net/burp/communitydownload

SensePost. (s.f.). Objection - Runtime Mobile Exploration. Recuperado el 19 de abril de 2024, de https://github.com/sensepost/objection

Honeypots como fuente de inteligencia de ciberamenazas

Paula Venosa [2], Patricia Bazán [2], Sofía Martin [2], Nicolás del Rio[1], Pablo Maddalena Kreff[1], Pablo Gagliardi[1], Francisco Javier Díaz[2]

UNLP, Facultad de Informática [1], LINTI UNLP, Facultad de Informática[2]

pvenosa@info.unlp.edu.ar, pbaz@info.unlp.edu.ar, smartin@linti.unlp.edu.ar, ndelrio@info.unlp.edu.ar, pkreff@linti.unlp.edu.ar, pablogagliardi91@gmail.com, jdiaz@info.unlp.edu.ar

RESUMEN

Los honyepots son herramientas que permiten detectar ataques informáticos y obtener información de los mismos, actuando como señuelos en un entorno controlado. La información recopilada por un honeypot, además de ser útil a las organizaciones, puede resultar una base conceptual para el proceso de enseñanza-aprendizaje en el área de la ciberseguridad.

El conjunto de estas herramientas, los honeypots, configurados en red, determinan una honeynet, y permiten recrear un escenario completo simulando diversos tipos de nodos de la red (sistemas en producción, servicios, servidores), ejecutándose en diversas infraestructuras entre las que también se incluyen elementos de IoT (Internet of Things).

En este trabajo se presenta un proyecto de investigación en torno a los honeypots como fuente de inteligencia en ciberamenazas, que combina el estudio de vulnerabilidades en IoT y los nuevos desafíos de seguridad que impone a las organizaciones, así como la aplicación de técnicas de ciencia de datos para el análisis de grandes volúmenes de información.

CONTEXTO

El proyecto se inició con el objetivo de establecer las metodologías necesarias para la implementación de una honeynet en una o más organizaciones, cuyo monitoreo y control se encuentre centralizado, pero la detección sea distribuida. Esto permitirá a las organizaciones tomar decisiones proactivas en base a las detecciones realizadas, utilizando técnicas de introspección de tráfico con el fin proteger las infraestructuras críticas en base a políticas definidas y mantenerse “invisibles” para los atacantes.

Además, el proyecto se propone consolidar un grupo de investigación especializado en ataques informáticos a nivel regional y, mediante el uso de honeypots, estudiar nuevas técnicas y tácticas empleadas en estos ataques.

Las líneas de acción definidas en este marco estuvieron orientadas a identificar distintas alternativas para implementar honeypots, elegir alguna/s e iniciar de ese modo la recolección de datos y su análisis para la detección de ataques.

Asimismo, se buscó integrar los honeypots, como fuente de información, con una herramienta de Cyber Threat Intelligence (CTI). La plataforma de CTI permite la recopilación, análisis y compartición de datos, recabados por los honeypots, mejorando así la capacidad de respuesta y prevención ante posibles ataques que puedan recibir las organizaciones con las que se comparte la información.

Finalmente, se está comenzando a abrir una línea de trabajo en torno a honeypots para IoT, así como la aplicación de la ciencia de datos -

una disciplina que combina estadística, aprendizaje automático y procesamiento de grandes volúmenes de información -, a los honeypots y así detectar patrones de ataque, predecir amenazas emergentes y optimizar la toma de decisiones en ciberseguridad.

1- INTRODUCCIÓN

Internet ha revolucionado la forma en que nos comunicamos y compartimos información. Sin embargo, con esta revolución también han surgido nuevos desafíos en términos de ciberseguridad. Aquí es donde entra en juego la importancia de proteger los sistemas, redes y aplicaciones frente a ciberataques.

A medida que los ataques se vuelven más sofisticados, la necesidad de una defensa proactiva se vuelve cada vez más crítica. En este sentido, la inteligencia de amenazas juega un papel crucial. Esta disciplina implica la recopilación y análisis de información sobre amenazas potenciales para ayudar a las organizaciones a prevenir o mitigar los ciberataques.

En este contexto, herramientas como MISP (Malware Information Sharing Platform) pueden ser extremadamente útiles. MISP es una plataforma de código abierto para compartir, almacenar y correlacionar indicadores de compromisos de ataques cibernéticos. Permite a las organizaciones compartir información de amenazas de manera eficiente y ayuda a mejorar la postura de seguridad general de una organización [5].

La recolección y análisis de información para detectar ataques de ciberseguridad plantea un desafío, ya que la detección de comportamiento malicioso y ataques mediante el análisis de tráfico de red sigue presentando dificultades a los responsables de monitorear la seguridad de la red y de gestionar los incidentes de seguridad. Aunque existen varios mecanismos de detección bien conocidos para separar con precisión los comportamientos maliciosos de los normales, todavía es extremadamente difícil contar con sistemas de detección eficientes [2] [4].

Los principales obstáculos para una buena detección de malware y ataques mediante el análisis de tráfico de red son cuatro: 1- que el tráfico normal es extremadamente complejo, diverso y cambiante, 2 - las acciones maliciosas cambian continuamente, adaptándose, migrando y ocultándose como tráfico normal, 3- la cantidad de datos para analizar es enorme y la capacidad de cómputo disponible para el análisis de datos es finita, lo que obliga a los analistas a descartar datos y 4- la detección debe ocurrir casi en tiempo real, o incluso antes que el ataque ocurra, para ser de alguna utilidad.

Los honeypots son herramientas que permiten el estudio de ataques de ciberseguridad en entornos controlados, con el objetivo de detectar el ataque y obtener información tanto sobre el ataque como sobre el atacante, con un nivel de detalle que otras herramientas no pueden proporcionar.

El conjunto de honeypots, configurados en red, determinan una honeynet, y permiten recrear un escenario completo simulando diversos tipos de nodos de la red (sistemas en producción, servicios, servidores, dispositivos inteligentes).

El objetivo de las honeynets es atraer atacantes y recopilar información sobre sus tácticas y herramientas.

Un honeypot puede ser de alta o baja interacción. Un honeypot de alta interacción proporciona al atacante un sistema operativo real con el que interactuar, donde nada es simulado o restringido [4]. Los honeypots de baja interacción emulan servicios específicos, proporcionando datos limitados pero útiles sin

el riesgo de comprometer la seguridad del sistema anfitrión [1].

Honeypots e IoT

Con el crecimiento de la Internet de las cosas y la proliferación de dispositivos de IoT, es fundamental analizar las tendencias de los ataques que tienen a éstos como objetivo.

Según un informe de Forrester Consulting[1], el 69% de las empresas tienen más dispositivos IoT en sus redes que computadoras, el 84% de los profesionales de la seguridad creen que los dispositivos IoT son más vulnerables que las computadoras, el 67% de las empresas han experimentado un incidente de seguridad de IoT, solo el 16% de los gerentes de seguridad empresarial dicen que tienen una visibilidad adecuada de los dispositivos IoT en sus entornos y el 93% de las empresas planean aumentar su gasto en seguridad para IoT y dispositivos no administrados.

La frecuencia de los ataques es tan alta que un dispositivo IoT tarda sólo cinco minutos en enfrentarse a ataques después de conectarse a Internet, según este estudio de NETSCOUT Threat Intelligence.

De acuerdo al informe de F-Secure de 2019[2], en el primer semestre de 2019 se registraron 2,900 millones de ataques a dispositivos de IoT en honeypots en todo el mundo. Eso es un aumento del 300% en comparación con el segundo semestre de 2018. Las versiones de Mirai botnet siguen dominando el panorama de IoT y son un factor importante detrás del aumento de los ataques DDoS.

Estos aspectos requieren un nuevo enfoque convergente de la seguridad. Junto con las cuestiones tradicionales de la seguridad escalonada, algunos aspectos adicionales entrarían en escena en el caso de los sistemas de IoT que combinan TI (Technology Information) y OT (Operational Technology) bajo un solo paraguas[3].

Un OT es el hardware y software que detecta o aplica un cambio mediante el monitoreo y/o control de dispositivos físicos, procesos y eventos dentro de una organización.

Según predicciones de Gartner[4], el impacto financiero de los ataques al sistema físico cibernético superará los 50,000 millones de dólares para 2023.

El análisis presentado condujo a la búsqueda en el repositorio Elsevier Scopus[5] que indexa las principales conferencias y revistas del área e incluye también las principales publicaciones de IEEE y ACM.

El resultado de la búsqueda arrojó como resultado la cantidad de publicaciones mostradas en la Tabla 1. Se observa un incremento en el número de trabajos en todos los tipos de búsquedas realizadas, así como un área de vacancia al combinar “honeynet” con “network traffic analysis”.

Tabla 1 - Cantidad de publicaciones por año y palabra clave

	Año de Publicación	2023	2022	2021	2020	2019	2018	Total
Cadena de búsqueda	honeynet & cybersecurity attack	1	5	4	1	6	2	19
	honeypot & atacck detection	13	60	50	41	52	44	260
	honeypot & IoT	4	29	25	34	27	14	133
	network traffic monitoring & atacck detection	20	167	145	94	111	100	637
	honeynet & network traffic analysis	0	0	1	1	2	1	5
								1054

[1] https://www.armis.com/analyst-reports/state-of-enterprise-iot-security-in-north-america-unmanaged-and-unsecured/

[2] https://blog-assets.f-secure.com/wp-content/uploads/2019/09/12093807/2019_attack_landscape_report.pdf

[3] https://www.capgemini.com/mx-es/insights/expert-perspectives/un-enfoque-convergente-de-la-ciberseguridad-de-iot/

[4] https://www.gartner.com/en/newsroom/press-releases/2020-09-01-gartner-predicts-75--of-ceos-will-be-personally-liabl

[5] https://www.elsevier.com/solutions/scopus

2- LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Las principales líneas de trabajo en las que el grupo de investigación desarrolla sus actividades y consolida su formación, resultan de la continuación del trabajo realizado por doctorandos y maestrandos en torno a estrategias de ciberseguridad distribuida [6] y el análisis de técnicas de ataque y comportamientos maliciosos [2].

A partir de estas líneas se plantea su continuidad en torno a dispositivos de IoT, implementación de honeypots y el análisis de grandes volúmenes de datos, identificación de patrones y visualización de resultados.

3- RESULTADOS ESPERADOS/OBTENIDOS

Los objetivos generales de este proyecto apuntaron a consolidar un grupo de trabajo que se especialice en la investigación de ataques informáticos a nivel regional, apuntando al más alto nivel y excelencia, dado que mediante el uso de honeypots, el estudio de nuevas técnicas y tácticas de ataque es posible.

Al trabajar en coordinación con CERTUNLP[6], el CSIRT Académico de la Universidad Nacional de La Plata, se cuenta con un ámbito de aplicación de lo investigado y de las pruebas realizadas, lo cual permite mejorar los servicios de ciberseguridad que se prestan a la UNLP.

Uno de los resultados obtenidos en el primer año de proyecto, está vinculado con la implementación de un honeypot con TPOT[7], completando su instalación en un entorno auxiliar no distribuido, así como de una herramienta de Inteligencia de Ciberamenazas y una herramienta SOAR (Seguridad, organización, automatización y respuesta). Se logró interactuar exitosamente con un honeypot y se estableció un mecanismo de interacción posible. La Figura 1 muestra cómo podría ser la interacción desde un atacante hasta una herramienta de CTI.

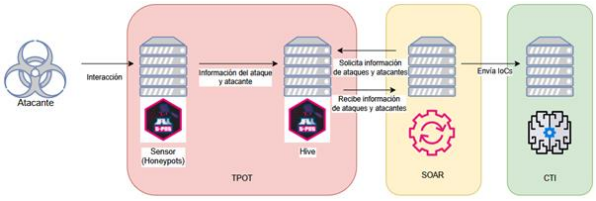


Figura 1 - Modelo de interacción entre un atacante y una herramienta de CTI

Por otra parte, las líneas de investigación del proyecto permitieron producir dos propuestas (una de grado y otra de postgrado).

Por un lado, la Práctica Profesional Supervisada, para la carrera Analista en Tecnologías de la Información y la Comunicación de la Facultad de Informática de la UNLP, de Pablo Gagliardi, bajo la tutoría de Nicolás del Río, cuyos objetivos incluyen comprender qué es un honeypot y los servicios que en él se simulan, investigar sobre herramientas de monitoreo de honeypots y participar en la implementación de sensores honeypot y el monitoreo de los mismos.

Por otro lado, la propuesta de Trabajo Final Integrador para la Especialización en Redes y Seguridad de la Facultad de Informática de la UNLP, de Pablo Germán Maddalena Kreff bajo la dirección de Paula Venosa y la codirección de Patricia Bazán, titulada “Gossip Hive: Honeypots como fuente de Inteligencia de Ciberamenazas”. El objetivo general de este Trabajo Integrador Final es relacionar a los honeypots como fuente de información con una herramienta de Cyber Threat Intelligence (CTI). Una plataforma de CTI permite la recopilación, análisis y compartición de datos, recabados por los honeypots, mejorando así la

[6] CERTUNLP: www.cert.unlp.edu.ar

[7] https://github.com/telekom-security/tpotce/blob/master/CITATION.cff

capacidad de respuesta y prevención ante posibles ataques que puedan recibir las organizaciones con las que se comparte la información.

4- FORMACIÓN DE RECURSOS HUMANOS

En esta línea de investigación trabaja un grupo de docentes/investigadores del LINTI (Laboratorio de Investigación en Nuevas Tecnologías Informáticas) de la Facultad de Informática de la UNLP (Universidad Nacional de La Plata), dando continuidad al trabajo en ciberseguridad iniciado hace muchos años en el Laboratorio [7].

En este área, el grupo ha realizado algunas experiencias previas en lo que refiere a plataformas de CTI, en particular en el desarrollo del Trabajo Final de carrera de Pablo Maddalena Kreff [5] donde se investigó al respecto y se colaboró en la puesta en producción de una red colaborativa para compartir indicadores de compromiso a través de la plataforma MISP[5], proyecto que se lleva adelante en la subcomisión de ciberseguridad del CIN[8] en coordinación con la ARIU[9], en la cual dos autores de este trabajo tienen una participación activa.

El trabajo de investigación presentado se enmarca en el proyecto de I+D+i "Redes de Honeypots: alcances, implementaciones y utilidad en la ciberseguridad de las organizaciones" de la Facultad de Informática de la UNLP llevado a cabo durante el año 2024.

5- BIBLIOGRAFÍA

1. Banerjee, M. (2021). Detection and behavioral analysis of botnets using honeynets and classification techniques. Distributed Denial of Service Attacks, pp. 131-158. De Gruyter. DOI: 10.1515/9783110619751-006.
2. Venosa, P. (2021). Detección de ataques de seguridad en redes usando técnicas de ensembling (Doctoral dissertation, Universidad Nacional de La Plata). http://sedici.unlp.edu.ar/handle/10915/120856. https://doi.org/10.35537/10915/120856
3. Mokube, I., & Adams, M. (2007, March). Honeypots: concepts, approaches, and challenges. In Proceedings of the 45th annual southeast regional conference (pp. 321-326).
4. Maddalena Kreff P., Gagliardi P., Bazán P., Venosa P., del Río N., Martin S., Bogado J (2024). Honeypots: análisis de implementaciones para ciberseguridad de una red organizacional. Argentina. La Plata 2024. CACIC 2024. http://sedici.unlp.edu.ar/handle/10915/172755. Pág.1289-1293.
5. Maddalena Kreff P. Malware Information Sharing Platform y su integración a CERTUNLP. Tesina de Grado. La Plata 2024. https://sedici.unlp.edu.ar/handle/10915/163504
6. Gallardo Urbini I. Estrategia de Ciberseguridad distribuida, aplicando el concepto de Operación de Inteligencia. Tesis Doctoral. La Plata 2023. http://sedici.unlp.edu.ar/handle/10915/152594
7. Diaz J, Venosa P, Macia N. Investigación en Ciberseguridad en un año de pandemia. WICC 2021. https://sedici.unlp.edu.ar/handle/10915/119490

[8] CIN: Comisión de Ciberseguridad del CIN

[9] ARIU: Asociación Redes de Interconexión Universitaria

Inteligencia Artificial aplicada a la Detección de Información de Identificación Personal (PII) en Argentina

Zea Cardenas Milagros Ninfa, Barrionuevo Ileana Maricel, Gibellini Fabian A., Parisi German N., Ruhl Analia L., Notreni Juliana M., Ciceri Leonardo R., Auquer Marcelo J., Bertola Federico, Sanchez Balzaretti Ignacio J.

GISSIC - Grupo de Investigación en Seguridad de Sistemas de Información y Ciberseguridad. / Laboratorio de Sistemas / Dpto. de Ingeniería en Sistemas / Universidad Tecnológica Nacional / Facultad Regional Córdoba
Cruz Roja S/N, 5016

milyzc@gmail.com, ilebarrionuevo@gmail.com, fabiangibellini@gmail.com, germannparisi@gmail.com, analialorenaruhl@gmail.com, julinotreni@gmail.com, leonardorciceri@gmail.com, marcelo.auquer@gmail.com, fedebertola@gmail.com, ignaciojsb@gmail.com

Resumen

La creciente relevancia de la ciberseguridad a nivel global ha puesto en evidencia la necesidad de proteger la Información de Identificación Personal (PII) dentro de las organizaciones. Este tipo de dato, regulado en Argentina bajo la Ley 25.326[1], suele estar presente en documentos organizacionales de manera dispersa, lo que complica su identificación y resguardo. Frente a esta problemática, el presente trabajo propone el desarrollo de un modelo de Inteligencia Artificial (IA) capaz de detectar automáticamente PII en documentos organizacionales, adaptándose a las regulaciones de privacidad de datos de la ley argentina vigente.

La problemática identificada es encontrar un modelo que logre garantizar la trazabilidad de los datos sensibles y minimizar los tiempos de respuesta ante incidentes de fuga de información. Adicionalmente, se desarrollará una herramienta de benchmarking para evaluar la precisión y eficiencia del modelo, permitiendo su validación en entornos reales. Este proyecto no solo contribuye a reforzar la seguridad de la información, sino que también promueve la transferencia de conocimiento a la academia, la industria y el sector público, fomentando la integración de soluciones basadas en IA en el ámbito de la seguridad informática.

Palabras clave: PII, ciberseguridad, seguridad, aprendizaje automático, IA, datos sensibles, protección de datos.

Contexto

En un mundo cada vez más digitalizado, la ciberseguridad se posiciona como una prioridad esencial para las organizaciones, destacándose como uno de los principales riesgos económicos y sociales a nivel global. Según el Foro Económico Mundial, la inseguridad digital ocupa actualmente el cuarto lugar entre los riesgos más significativos y se proyecta que, para 2034, mantendrá su relevancia como uno de los ocho principales desafíos a enfrentar [2].

En este contexto, la información de identificación personal (PII) se establece como un activo crítico, la cual puede generar graves consecuencias legales, económicas y reputacionales. La PII incluye datos como nombres, direcciones, números de identificación personal y otros elementos que, si se encuentran

expuestos, pueden ser utilizados para fraudes, suplantación de identidad o espionaje [3]. A nivel internacional existen reglamentos con el fin de proteger información privada y reglamentos de protección de datos como HIPAA [4], FCRA [5], FERPA [6], GLBA [7], ECPA [8], COPPA [9] y VPPA [10]. En Argentina, la Ley 25.326 regula la protección de datos personales, estableciendo obligaciones para las organizaciones respecto al tratamiento y resguardo de este tipo de información [11]. Sin embargo, la identificación de PII en documentos organizacionales es un desafío debido a su presencia en formatos variados y contextos inesperados, incluyendo datos indirectos que pueden ser inferidos a partir de combinaciones de información [12].

Por otro lado, las violaciones de datos son cada vez más comunes y abarcan desde grandes empresas tecnológicas hasta organismos públicos. Ejemplos recientes, como la filtración de 700 millones de datos de LinkedIn o las brechas de seguridad en Microsoft, han marcado la importancia de contar con herramientas capaces de identificar y proteger esta información sensible [13][14].

Frente a esta problemática, resulta fundamental diseñar soluciones tecnológicas que permitan identificar y proteger la PII de manera eficiente. Este proyecto busca proponer un modelo de Inteligencia Artificial que, además de cumplir con los marcos normativos de la Argentina, tiene como objetivo principal garantizar la seguridad de los datos y optimizar la capacidad de las organizaciones para responder ante eventuales fugas de información. Al focalizarse en datos en español y regulaciones locales, este modelo lo que busca es superar las limitaciones de las herramientas actuales, que suelen estar orientadas a normativas y contextos internacionales [15].

Introducción

En esta era digital, la cantidad de datos que las organizaciones generan y almacenan crece de forma exponencial. Estos datos incluyen desde información operativa hasta datos personales, que se encuentran entre los más sensibles y vulnerables. La Información de Identificación Personal (PII, por sus siglas en inglés) se define como cualquier representación de datos que pueda identificar directa o indirectamente a una persona. Ejemplos de PII incluye nombres, direcciones, números de identificación personal, información financiera e incluso datos contextuales que permiten inferir información sensible, como condiciones médicas o historiales laborales [3][12].

El manejo y protección de PII se han convertido en un tema muy popular en ciberseguridad, ya que su exposición no solo puede provocar graves consecuencias legales y económicas, sino también daños en la reputación que son irreparables para las organizaciones. Este desafío se magnifica en un contexto global donde las violaciones de datos son cada vez más frecuentes y de mayor alcance.

La Ley 25.326 de Argentina establece un marco normativo para la protección de los datos personales, obligando a las organizaciones la responsabilidad de adoptar medidas para garantizar la seguridad de estos datos. Sin embargo, cumplir con estas normativas no es tarea sencilla, especialmente considerando que la PII puede estar dispersa en documentos no estructurados y almacenada en formatos variados. Este escenario es por el cual se plantea la necesidad de desarrollar herramientas tecnológicas capaces de identificar y gestionar automáticamente estos datos, asegurando su protección en todo momento [11].

El presente surge como respuesta a esta problemática, proponiendo el desarrollo de un modelo de Inteligencia Artificial (IA) orientado a la detección automática de PII en documentos organizacionales. Este modelo se basará en técnicas avanzadas de aprendizaje automático y procesamiento del lenguaje natural, adaptándose

específicamente al idioma español y a las regulaciones locales. Se está contemplando no sólo la identificación de datos explícitos, como nombres o direcciones, sino también la detección de datos que podrían clasificarse como sensibles según el contexto de la organización.
Además del desarrollo del modelo de IA, se plantea la creación de una herramienta de benchmarking que permita evaluar de manera integral el rendimiento del modelo, analizando métricas como precisión, sensibilidad y capacidad de generalización. Este enfoque asegura que el modelo sea eficaz y que pueda ser aplicable en diversos entornos, como organizaciones locales, instituciones académicas o gubernamentales.
Este proyecto no se limita a resolver el problema inmediato de detectar PII, sino que también busca sentar bases sólidas para manejar de forma automatizada este tipo de datos en Argentina. La idea es que los resultados puedan aplicarse tanto en empresas como en instituciones académicas o del sector público. Además, la herramienta está pensada para ayudar a cumplir con las normativas locales y disminuir los riesgos de fuga de datos.

Líneas de Investigación y Desarrollo

Esta investigación se basa en los conocimientos más avanzados y contemporáneos en el campo de la ciberseguridad, con un enfoque específico en la protección de Información de Identificación Personal (PII). La línea de investigación principal será seguridad informática, pero integra además de forma transversal técnicas y algoritmos provenientes del ámbito de Machine Learning, convirtiéndose en un proyecto multitemático que combina disciplinas clave para abordar esta problemática.
El enfoque metodológico es empírico, basado en el método científico para la recolección y análisis de datos necesarios para el desarrollo del modelo de Inteligencia Artificial (IA) destinado a la detección automatizada de PII. Este modelo busca garantizar la trazabilidad y protección de datos sensibles, cumpliendo con las normativas locales y adaptándose a las particularidades del idioma español y los entornos organizacionales argentinos.
Este trabajo se desarrolla bajo los principios de Investigación, Desarrollo e Innovación (I+D+i), integrando una visión estratégica que combina la teoría y la práctica con un enfoque orientado a resultados. Las actividades se estructuran abarcando: síntesis y teoría; exploración, hipótesis y clarificación; diseño, desarrollo y prueba; implementación; y mejora continua basada en la evaluación de la eficacia del producto final.

Resultados y Objetivos

El objetivo principal del proyecto es desarrollar un modelo de Inteligencia Artificial (IA) que permita la detección automática de Información de Identificación Personal (PII) en documentos organizacionales, considerando las regulaciones de la Ley 25.326 de Argentina. Este modelo estará diseñado para proporcionar trazabilidad de los datos sensibles, reducir los tiempos de respuesta ante incidentes de fuga de información y colaborar en la toma de decisiones estratégicas para prevenir futuros desastres.
Además, el proyecto contempla el desarrollo de una herramienta de benchmarking que encapsule el proceso de evaluación del modelo de IA, asegurando su eficacia y precisión en entornos reales. Estas herramientas están orientadas a fortalecer la seguridad de la información en organizaciones del ámbito académico, industrial y gubernamental.
Para alcanzar estos objetivos, el proyecto se estructura en las siguientes metas:

- Colaborar en la definición de estrategias de contingencia, partiendo del conocimiento de qué datos de PII han sido fugados de una organización.
- Determinar una clasificación y severidad (criticidad) de datos de PII para evaluar su

importancia y sensibilidad dentro de una organización, basado este proceso en pautas brindadas por NIST e ISO27001 [16][17].

- Identificar algoritmos y modelos de IA existentes para la detección de PII, evaluando sus tasas de éxito y aplicabilidad bajo las regulaciones locales.
- Diseñar un modelo de IA que detecte PII en documentos organizacionales, adaptado al idioma español y ajustado a las regulaciones de la Ley 25.326 de Argentina.
- Desarrollar una herramienta de benchmarking, que permita probar y validar el modelo de IA, estableciendo métricas como precisión, sensibilidad y usabilidad.
- Aplicar métodos de entrenamiento para el modelo desarrollado para la detección de PII en documentos de una organización.
- Posterior a la implementación técnica de cada meta anteriormente mencionada, se plantea como posibilidad ampliar el desarrollo para detectar PII en sistemas y aplicaciones, a través del tráfico web.

Con estos resultados, se espera mejorar notablemente la gestión de datos sensibles en las organizaciones, ofreciendo soluciones prácticas que faciliten la detección y el manejo de PII. Se busca aportar una herramienta útil y efectiva para abordar las necesidades concretas del contexto local, ayudando a las organizaciones a manejar su información de forma más segura y eficiente.

Formación de Recursos Humanos

Los involucrados en el proyecto son docentes e investigadores de la Universidad Tecnológica Nacional - Facultad Regional Córdoba (UTN - FRC) y también integrantes del Grupo de Investigación en Seguridad de Sistemas de Información y Ciberseguridad [18].

También participan alumnos becarios de la carrera de Ingeniería en Sistemas de Información. En lo cual los estudiantes pueden iniciar su formación en investigación científica y tecnológica, y ahondar en temas como la seguridad de la información y el desarrollo de software.

Se organizan charlas, talleres y conferencias tanto a nivel provincial como nacional, con el objetivo de compartir los resultados del proyecto.

Referencias

[1] Protección de datos personales. https://www.argentina.gob.ar/normativa/nacional/ley-25326-64790/actualizacion, última visita: 21/02/2025.

[2] World Economic Forum. https://www3.weforum.org/docs/WEF_The_Global_Risks_Report_2024.pdf.

[3] Guidance on the Protection of Personal Identifiable Information. Departament of labor U.S. https://www.dol.gov/general/ppii#:~:text=Personal%20Identifiable%20Information%20(PII)%, última visita 21/02/2024.

[4] HIPAA for Individuals. https://www.hhs.gov/hipaa/for-individuals/index.html, última visita 27/04/2024.

[5] Fair Credit Reporting Act. https://www.ftc.gov/legal-library/browse/statutes/fair-credit-reporting-act, última visita 27/04/2024.

[6] Family Educational Rights and Privacy Act. https://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html, última visita 27/04/2024.

[7] Gramm-Leach-Bliley Act. https://www.ftc.gov/business-guidance/privacy-security/gramm-leach-bliley-act, última visita 27/04/2024.

[8] Electronic Communications Privacy Act of 1986. https://bja.ojp.gov/program/it/privacy-civil-liberties/authorities/statutes/1285, última visita 27/04/2024.

[9] Children's Online Privacy Protection Rule ("COPPA"). https://www.ftc.gov/legal-

library/browse/rules/childrens-online-privacy-protection-rule-coppa, última visita 27/04/2024.
[10] Video Privacy Protection Act. https://en.wikipedia.org/wiki/Video_Privacy_Protection_Act, última visita 27/04/2024.
[11] Protección de los datos personales. Decreto 1558/2001. https://servicios.infoleg.gob.ar/infolegInternet/anexos/70000-74999/70368/texact.htm.
[12] Defining and detecting Personally Identifiable Information (PII) in log data. Splunk Platform. https://lantern.splunk.com/Splunk_Platform/UCE/Security/Compliance/Defining_and_detecting_Personally_Identifiable_Information_(PII)_in_log_data.
[13] Hackers leak LinkedIn 700 million data scrape. Catalin Cimpanu. Septiembre 2021. https://therecord.media/hackers-leak-linkedin-700-million-data-scrape, última visita 28/04/2024.
[14] Microsoft Data Breaches: Full Timeline Through 2024. M. X. Heiligenstein. https://firewalltimes.com/microsoft-data-breach-timeline/, última visita 28/04/2024.
[15] Silva, P., Gonçalves, C., Godinho, C., Antunes, N., & Curado, M. (2020, July). Using nlp and machine learning to detect data privacy violations. In IEEE INFOCOM 2020-IEEE conference on computer communications workshops (INFOCOM WKSHPS) (pp. 972-977).
[16] NIST SP 800-122, Guide to Protecting the Confidentiality of Personally Identifiable Information (PII). https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-122.pdf, última visita 28/04/2024.
[17] Qué es la ISO 27001, https://www.normaiso27001.es/, última visita 21/03/2025.
[18] Grupo de Investigación en Seguridad de Sistemas de Información y Ciberseguridad. https://gissic.frc.utn.edu.ar/

OTP-Vote: Mejoras en Seguridad y Verificabilidad para el Voto Electrónico

Silvia Bast[1] Germán Montejano[2] MarioBerón[2]

[1]Departamento de Matemática
Facultad de Ciencias Exactas y Naturales
Universidad Nacional de La Pampa
Av. Uruguay 151 – (6300) Santa Rosa – La Pampa – Argentina
Tel.: +54-2954-425166– Int. 28
silviabast@exactas.unlpam.edu.ar – web: http://exactas.unlpam.edu.ar/

[2]Departamento de Informática
Facultad de Ciencias Físico Matemáticas y Naturales
Universidad Nacional de San Luis
Ejército de los Andes 950 – (5700) San Luis – San Luis – Argentina
Tel.: +54-2652-424027 – Int. 251
[gmonte, mberon]@unsl.edu.ar – web: http://www.unsl.edu.ar

RESUMEN

Las elecciones representan un componente esencial de la democracia moderna por su impacto directo en la conformación de los gobiernos y la vida ciudadana. El presente trabajo se centra en el desarrollo y mejora del sistema de voto electrónico OTP-Vote, abordando uno de los desafíos más significativos: la implementación de sistemas de votación electrónica confiables y seguros. Partiendo de un modelo inicial de datos, la investigación ha profundizado en aspectos críticos para fortalecer la seguridad y confiabilidad del sistema. Los avances se han enfocado en el desarrollo de atributos de control y encriptación, el refinamiento de protocolos antifraude, y la implementación de mecanismos de verificabilidad End to End. Un aspecto fundamental ha sido el diseño de un modelo para automatizar la configuración de parámetros y la generación de estructuras de datos. La investigación aborda directamente la principal barrera para la adopción del voto electrónico: la desconfianza ciudadana basada en experiencias previas fallidas. El proyecto busca superar estas resistencias mediante la construcción de un sistema que no solo sea técnicamente robusto, sino que también permita una auditoría en las etapas del proceso electoral, garantizando tanto el secreto del voto como la integridad del proceso.

Palabras clave: Sistemas de Voto Electrónico, Anonimato, Transparencia, Auditoría, One Time Pad, Verificabilidad End to End.

CONTEXTO

Los avances presentados se enmarcan en el Proyecto de Investigación denominado "Especificación Integral del Sistema OTP-Vote Orientada a su Implementación" (Resolución N° 55/22 del Consejo Directivo FCEyN – UNLPam) que es la continuación de la línea de investigación “Desarrollo de un Modelo de Voto Electrónico basado en Criptografía One Time Pad” del proyecto "Aspectos de Seguridad en Proyectos de

Software" (Resolución N° 488/14 del Consejo Directivo de FCEyN – UNLPam).
El proyecto es dirigido por el Doctor Germán Montejano (Universidad Nacional de San Luis) y por el Doctor Mario Berón (Universidad Nacional de San Luis) y se desarrolla en el ámbito de la Facultad de Ciencias Exactas y Naturales de la Universidad Nacional de La Pampa.

1. INTRODUCCIÓN

En las sociedades democráticas modernas, los procesos electorales constituyen el mecanismo fundamental para transformar la voluntad ciudadana en decisiones de gobierno. Esta transformación requiere sistemas que no solo aseguren la integridad del proceso, sino que también generen confianza en el electorado. La evolución tecnológica actual sugiere la posibilidad de digitalizar estos procesos, sin embargo, la naturaleza única del voto como expresión democrática presenta desafíos particulares.
El contraste entre los sistemas de voto electrónico y otras aplicaciones digitales resulta significativo. Mientras que la mayoría de las transacciones digitales buscan mantener registros detallados que vinculen usuarios con operaciones, el voto electrónico debe garantizar precisamente lo opuesto: la imposibilidad de establecer conexiones entre el votante y su elección. Esta característica fundamental, junto con experiencias previas poco exitosas en diversos contextos, ha generado un debate intenso sobre la viabilidad y confiabilidad de estos sistemas.
El sistema manual actual, si bien es ampliamente aceptado, presenta vulnerabilidades significativas que incluyen prácticas irregulares como la sustracción de boletas, la manipulación de urnas y la alteración de actas. Además, enfrenta desafíos logísticos considerables relacionados con los costos de impresión de boletas (principalmente en aquellos lugares en los que no se hace uso de boleta única) y la lentitud en el proceso de escrutinio, especialmente en distritos con gran cantidad de votantes y múltiples candidatos.
En este contexto, el modelo OTP-Vote propone una base teórica sólida que requiere ser expandida y especificada en detalle para su implementación práctica. El presente artículo aborda los avances realizados en la optimización de este modelo, con especial énfasis en los aspectos de seguridad y verificabilidad que resultan críticos para su eventual adopción.
Para abordar este desafío, se parte de tres premisas fundamentales: i) los sistemas de voto electrónico pertenecen a la categoría de sistemas críticos, dada su relevancia en los procesos democráticos, ii) el sistema manual actualmente en uso, si bien ofrece prestaciones aceptables, presenta limitaciones significativas, y iii) el diseño y desarrollo de un sistema robusto y confiable representa un desafío complejo que requiere un enfoque integral.

Sistemas de Voto Electrónico

Según Hao y Ryan [1] desde el punto de vista matemático, un sistema de voto puede verse como una función que toma como entradas las preferencias de los electores y devuelve como salida el resultado de la elección.
Los sistemas de voto electrónico comparten ciertos requisitos básicos con otros sistemas tecnológicos, pero su naturaleza particular exige características adicionales específicas. Diversos investigadores como Epstein [2], Alam y Tamura [3] y van de Graaf [4] han propuesto conjuntos de requerimientos fundamentales para estos sistemas. Entre los más críticos se destaca el secreto del voto, que implica la imposibilidad de vincular al votante con su elección, y la elegibilidad, que garantiza que solo los votantes autorizados puedan participar.
La verificabilidad emerge como un requisito multifacético que abarca desde la confirmación individual del voto hasta la integridad del recuento final. Paralelamente, el sistema debe mantener un equilibrio entre la simplicidad de uso para el votante y la robustez necesaria para resistir intentos de manipulación. La auditabilidad y la inviolabilidad son esenciales

para prevenir el cómputo parcial prematuro y garantizar la transparencia del proceso.

Con la evolución de las investigaciones, han surgido requerimientos adicionales más sofisticados. Entre estos, destaca la verificabilidad "End to End" (VE2E), que según Ryan et al [5], tiene como objetivo garantizar a todas las partes involucradas que el resultado electoral refleje fielmente los votos emitidos de manera legítima. Para ello, la seguridad del proceso no debe depender únicamente del correcto funcionamiento de sus componentes, sino de la generación de evidencia inmutable a lo largo de la ejecución de la elección. Autores como Rabin y Rivest [6] y Awad y Leiss [7], Chaum et al [8], [9] y Benaloh et al [10], profundizan sobre esta característica.
Otros aspectos cruciales incluyen la independencia del software, definida por Rivest y Virza [11], y la necesidad de evidencia física que respalde la transparencia del proceso [1].
Un desafío particular emerge de la aparente contradicción entre algunos de estos requisitos, especialmente entre la verificabilidad y el secreto del voto, lo que añade complejidad al diseño e implementación de estos sistemas.

El Modelo OTP-Vote

En [12] se presentan las bases del modelo teórico OTP-Vote que se fundamenta en la siguiente premisa. En los sistemas de Voto Electrónico es necesario:
• Proteger la privacidad del votante indefinidamente, aún después de que termine el proceso eleccionario. El sustento de este requerimiento radica en que, si algún intruso obtiene una copia digital de los datos de la elección, podría intentar descifrarlos sin importar el tiempo que ese proceso demande.
• Proteger los datos relacionados con los votos mientras dure el proceso electoral, ya que posteriormente esos datos se hacen públicos y todos los electores tendrán acceso a los mismos.
OTP - Vote ofrece anonimato incondicional y seguridad computacional, que puede llevarse al nivel exigible, durante el proceso electoral.

El proceso consiste de tres etapas que pueden observarse en la Figura 1.

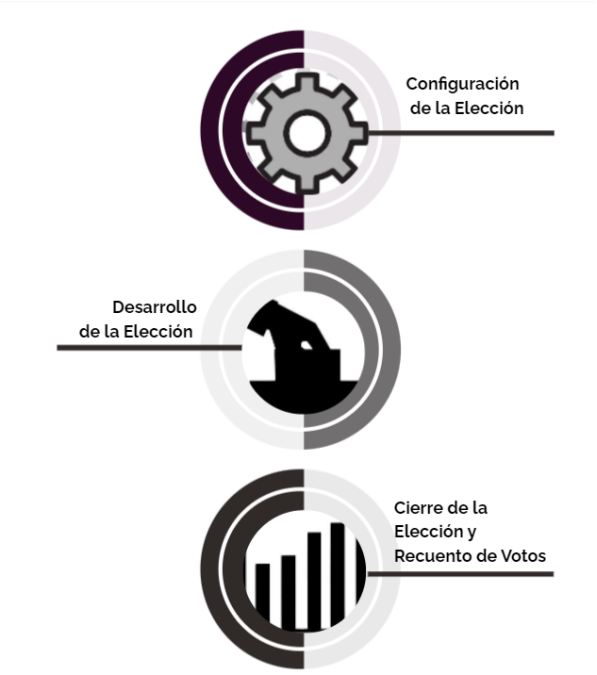


Figura 1. Etapas del Proceso Eleccionario en OTP- Vote

Para su funcionamiento OTP-Vote hace uso de Claves One Time Pad (OTP) [13], el esquema de almacenamiento Múltiples Canales Datos único (MCDU) [14] y la redundancia apropiada [15] en el almacenamiento de los datos.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Las líneas de investigación y desarrollo se orientan hacia el refinamiento integral del modelo OTP-Vote, con el objetivo de transformar sus bases teóricas en un sistema robusto e implementable. El proyecto avanza sobre dos ejes principales de desarrollo:

- El primer eje se centra en la optimización del modelo base, focalizándose en el fortalecimiento de las funcionalidades existentes. Esta línea incluye ajustes y mejoras que buscan incrementar el desempeño y la robustez del sistema, manteniendo sus

características fundamentales de seguridad y confiabilidad.

- El segundo eje aborda la incorporación de nuevos elementos al modelo, con especial énfasis en los mecanismos de auditoría y VE2E. Estas contribuciones resultan críticas para aumentar la transparencia del sistema sin comprometer el anonimato del votante, aspecto fundamental para generar confianza en el electorado.

Ambas líneas de desarrollo convergen en el objetivo común de demostrar la confiabilidad del sistema OTP-Vote, abordando de manera sistemática cada uno de los aspectos que requieren mayor precisión y especificación. Las propuestas desarrolladas en cada eje se complementan para construir un sistema que no solo sea técnicamente sólido, sino también prácticamente implementable.

3. RESULTADOS Y OBJETIVOS

Para alcanzar el objetivo general del proyecto, "*Ampliar la formalización del modelo OTP-Vote, fortaleciendo sus condiciones de seguridad y permitiendo su implementación como un sistema de voto electrónico robusto, verificable y transparente",* se avanzó en las siguientes acciones discriminadas por ejes.

- Primer eje: Estas mejoras están relacionadas con la optimización de las configuraciones y el aseguramiento de condiciones óptimas de seguridad en el sistema. Los logros específicos incluyen:
  - La identificación y clasificación de datos según su nivel de inmutabilidad, estableciendo protocolos específicos para el manejo de aquellos que requieren modificaciones controladas.
  - El refinamiento de la semántica de las tuplas incorporando elementos de control que optimizan la gestión de la información y garantizan la consistencia operativa.

  Estas mejoras resultan un aporte a la optimización de la configuración del sistema y la generación automática de las estructuras de datos a partir de la parametrización inicial provista por la Junta Electoral.

- Segundo eje. Nuevos aportes al modelo. Este aspecto se enfoca en desarrollar mecanismos que fortalezcan la confianza en el sistema. Los principales logros incluyen:
  - El desarrollo de una propuesta de auditoria del proceso eleccionario, que permite que las autoridades puedan verificar la correctitud del proceso sin comprometer la privacidad del votante.
  - La especificación y validación de una propuesta de VE2E que permite a los electores realizar el seguimiento de su voto mientras se preserva su anonimato.

  Estas contribuciones han resultado fundamentales para construir un sistema donde tanto votantes como autoridades dispongan de herramientas efectivas para verificar la integridad del proceso electoral.

4. FORMACIÓN DE RECURSOS HUMANOS

En lo que respecta a la formación de recursos humanos, Silvia Bast se encuentra finalizando la tesis titulada "Especificación Integral del Modelo OTP-Vote orientada a su implementación", con el objetivo de obtener el grado de Doctora en Ingeniería Informática por la Universidad Nacional de San Luis. Este trabajo ha sido formalmente registrado y aprobado mediante la Resolución de Inscripción y Aprobación de Plan de Tesis 408/21 emitida por el Decanato de la Facultad de Ciencias Físico-Matemáticas y Naturales de la mencionada universidad.

5. BIBLIOGRAFÍA

[1] F. Hao, P. Ryan. "Real-world electronic voting: Design, analysis and deployment." CRC Series in Security, Privacy and Trust. CRC Press, 2016. ISBN-13: 978-1498714693. ISBN-10: 1498714692. 2017.

[2] J. Epstein, "Electronic Voting" in Computer, vol. 40, no. 8, pp. 92-95, Aug 2007. doi: 10.1109/MC.2007.271.

[3] K. M. Rokibul Alam and S. Tamura, "Electronic voting - Scopes and limitations", International Conference on Informatics, Electronics & Vision (ICIEV), Dhaka, Bangladesh, pp. 525-529, . 2012. doi: 10.1109/ICIEV.2012.6317324.

[4] J.van de Graaf, C. Henrich and J. Müller-Quade, "Requirements for secure voting", Work Notes 2011.

[5] P. Ryan, S. Schneider, y V Teague. Prˆet `a voter—the evolution of the species. In Real-World Electronic Voting, pages 325–358. Auerbach Publications, 2016.

[6] M. Rabin and R. Rivest. Efficient end to end verifiable electronic voting employing split value representations. EVOTE 2014.

[7] M. Awad and E. Leiss. End-to-end cryptography: Spreading democracy. International

[8] D. Chaum, A. Essex, R. Carback, J. Clark, S. Popoveniuc, A. Sherman, and P. Vora. "Scantegrity: End-to-end voter verifiable optical-scan voting". IEEE Security & Privacy, 6(3):40–46, 2008.

[9] D. Chaum, R. Carback, J. Clark, A. Essex, S. Popoveniuc, R. Rivest, P. Ryan, E. Shen, A. Sherman and P. Vora. "Scantegrity II: End-to-end verifiability by voters of optical scan elections through confirmation codes". IEEE transactions on information forensics and security, 4(4), 611-627. 2009.

[10] J. Benaloh, R. Rivest, P. Ryan, P. Stark,V. Teague, & P. Vora,(2015). End-to-end verifiability. arXiv preprint arXiv:1504.03778.

[11] R. Rivest and M. Virza. Software independence revisited. In Real-World Electronic Voting, pages 19–34. Auerbach Publications, 2016.

[12] S. Bast. "Optimización de la integridad de datos en sistemas de e-voting". Master's thesis, Universidad Nacional de San Luis, Dic 2016.

[13] C. E. Shannon, "Communication theory of secrecy systems," in The Bell System Technical Journal, vol. 28, no. 4, pp. 656-715, Oct. 1949, doi: 10.1002/j.1538-7305.1949.tb00928.x..

[14] P. García, "Una Optimización para el Protocolo Non Interactive Dining Cryptographers" - Editorial Académica Española (https://www.eae-publishing.com/ - ISBN-13: 978-3-639-85270-7. ISBN-10: 3639852702. EAN: 9783639852707 – 2017.

[15] P. García, G. Montejano, S. Bast, E. Fritz, "Codificación de Sufragios con Detección de Colisiones en NIDC con Canales Paralelos de Slots" Congreso Nacional de Ingeniería en Informática / Sistemas de Información. CoNaIISI 2016.

PROTECCIÓN PROACTIVA DE LA IDENTIDAD DIGITAL DE USUARIOS EN WEB 2.0 Y WEB 3.0 MEDIANTE INTELIGENCIA ARTIFICIAL

Miguel Méndez-Garabetti[1,2,3], Eduardo Piray[3,4], Leonardo Requena[4], Ricardo González[3,4], Guillermo Romero Arregin[3,4], Federico Schwemler[4]

[1]*Facultad de Ciencias Sociales y Administrativas, Universidad del Aconcagua, Mendoza, Argentina.*

[2]*Free and Open Source Software/Hardware Research Laboratory (FOSSHLab), Argentina.*

[3]*Departamento de Sistemas, Universidad CAECE, Mar del Plata, Argentina.*

[4]*Universidad Siglo 21, Córdoba, Argentina.*

mmendezgarabetti@gmail.com, edupiray@gmail.com, lprequena@gmail.com, eliangonzalez045@gmail.com , gromeroarregin@gmail.com, schwemlerfederico@gmail.com

RESUMEN

Las aplicaciones web están permanentemente conectadas a redes de datos, y más que todo a la red Internet. Por lo tanto, las mismas están expuestas a ciberataques. Es clave efectuar una correcta Gestión de la Identidad y Acceso (Identity and Access Management IAM por su siglas en inglés), porque la identidad digital de los usuarios y los activos digitales deben estar protegidos, y no deben permitirse accesos no autorizados. En este ámbito, la inteligencia artificial integrada con el machine mearning (aprendizaje automático) y mediante el uso de redes neuronales artificiales, es posible administrar de forma automática y proactiva la seguridad en dichas aplicaciones. Es clave poder demostrar cómo detectar dinámicamente amenazas y cómo optimizar la gestión de identidad y acceso mediante inteligencia artificial en Web 2.0 y Web 3.0. Es importante integrar la inteligencia artificial, el aprendizaje automático y las redes neuronales para optimizar la gestión de Identidad y Acceso en aplicaciones web. Partiendo de las limitaciones de los sistemas de gestión tradicionales que no utilizan inteligencia artificial, se explorará cómo los algoritmos de machine learning con redes neuronales pueden mejorar la autenticación, la detección de anomalías y la autorización en tiempo real. Este enfoque busca avanzar hacia un modelo IAM autónomo, seguro y escalable. Se pretende seguir una metodología de estudio basada en un análisis cuantitativo. Por medio de estudios empíricos en ambientes de pruebas controlados que permitan evaluar la efectividad de los modelos de IAM generados.

Palabras claves: Gestión de Identidades y Accesos (IAM). Inteligencia artificial. Aprendizaje automático. Aprendizaje automático no supervisado. Ciberseguridad proactiva.

CONTEXTO

La presente línea de investigación consiste en una iniciativa encabezada por el FOSSHLab (Free and Open Source Software/Hardware Research Laboratory), institución que impulsa el desarrollo de proyectos centrados en tecnologías de software y hardware libre. Actualmente la línea cuenta con tres

proyectos de investigación llevados a cabo en universidades diferentes: 1) Universidad Siglo 21, en el marco de la carrera Licenciatura en Informática, 2) Universidad CAECE, en el marco de la Licenciatura en Sistemas, la Ingeniería en Sistemas y la Licenciatura en Gestión de Sistemas, 3) Universidad del Aconcagua. Cada proyecto está financiado individualmente por su respectiva universidad.

1. INTRODUCCIÓN

El término ciberseguridad se refiere a un conjunto de tecnologías, procesos y prácticas para proteger y defender redes, dispositivos, software y datos de ataques, daños o accesos no autorizados. La ciberseguridad se está volviendo compleja debido al crecimiento exponencial de dispositivos, sistemas y redes interconectados. Esto se ve exacerbado por los avances en la tecnología y la infraestructura digitales, lo que lleva a un crecimiento significativo de los ciberataques con graves consecuencias. Esta evolución está impulsando un aumento en la cantidad, la escala y el impacto de los ciberataques, y requiere la implementación de la ciberseguridad basada en inteligencia artificial para proporcionar una defensa dinámica contra los ciberataques en evolución. Las organizaciones de asesoramiento, como el Instituto Nacional de Estándares y Tecnologías (National Institute of Standards and Technologies NIST), también están fomentando el uso de enfoques más proactivos y adaptativos al pasar a evaluaciones en tiempo real, monitoreo continuo y análisis basados en datos para identificar, proteger, detectar, responder y catalogar ciberataques para prevenir futuros incidentes de seguridad [1].

La IAM depende cada vez más de la inteligencia artificial lo que permite a las empresas adoptar una respuesta mucho más detallada y adaptable a la autenticación y el control de acceso [2] [3]. Además, se requiere inteligencia artificial para el análisis de comportamiento de los usuarios y entidades, que se utiliza para detectar actividades sospechosas. La autenticación y el control de acceso son necesarios para la seguridad de los datos [2].

El uso de la inteligencia artificial por parte de los ciberdelincuentes para llevar a cabo ataques cada vez más automatizados y amplios, se está volviendo más común. La reacción obvia sería restringir el control de acceso en la mayor medida posible, obligando a los usuarios a autenticarse varias veces durante el día mediante MultiFactor Authentication (MFA) cada vez que necesiten acceder a sus aplicaciones [4]. Si bien un sistema IAM puede permitir implementar tales reglas de seguridad, no todos pueden aumentar o reducir dinámicamente los requisitos de autenticación necesarios [2].

El uso de enfoques de inteligencia artificial, específicamente técnicas de aprendizaje automático no supervisado (unsupervised machine learning), es muy prometedor para abordar la complejidad de las amenazas internas y los ataques relacionados con el robo de identidad. Al aprovechar el poder de la inteligencia artificial, las organizaciones pueden mejorar su capacidad para detectar y mitigar estas amenazas, salvaguardando así la información confidencial y los activos valiosos. La investigación y el desarrollo continuos en este campo son esenciales para mantenerse a la vanguardia del panorama en constante evolución de las actividades maliciosas, asegurando un futuro seguro y resiliente tanto para las personas como para las organizaciones [5].

Por otro lado, al aprovechar el poder de la inteligencia artificial, los sistemas de detección de amenazas pueden mejorar en gran medida la precisión y la eficiencia de los sistemas de gestión de identidad y acceso. Los sistemas IAM, a su vez, desempeñan un papel vital en la protección de las organizaciones contra el acceso no autorizado y las

actividades maliciosas. Con el apoyo de las tecnologías de inteligencia artificial, los sistemas IAM pueden analizar cantidades masivas de datos en tiempo real, lo que les permite detectar patrones o comportamientos anormales que pueden indicar una posible amenaza a la seguridad. Este enfoque proactivo para la detección de amenazas proporciona a las organizaciones la capacidad de responder rápidamente a los incidentes de seguridad y prevenir posibles infracciones antes de que ocurran [5].

La capacidad de los sistemas de inteligencia artificial de aprender y adaptarse sobre la marcha es fundamental para su capacidad de recopilar información y responder correctamente a las condiciones cambiantes. Esta característica mejora la estrategia de gestión de acceso inteligente al proporcionar una arquitectura más adaptativa. El término "control de acceso personalizado" se refiere al uso de inteligencia artificial para adaptar los niveles de acceso para usuarios individuales en función de criterios como sus responsabilidades, comportamientos pasados y niveles percibidos de riesgo. Esta técnica permite una gestión de permisos más profunda y un mayor conocimiento del contexto. El modelado predictivo de riesgos emplea inteligencia artificial para examinar tendencias y correlaciones para anticipar posibles riesgos de seguridad [6].

Los antecedentes mencionados plantean distintas problemáticas de ciberseguridad que pueden estudiarse en el presente proyecto.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

La línea de investigación se centra en la ciberseguridad tomando como eje la protección de la identidad del usuario en aplicaciones web. A continuación, se describe brevemente la línea de investigación.

1. Por un lado se trabaja con el objetivo de demostrar cómo detectar dinámicamente amenazas y cómo optimizar la gestión de identidad y acceso mediante inteligencia artificial en Web 2.0 y Web 3.0 para proteger la identidad digital y los accesos no autorizados. Se tiene como objetivo generar modelos de IAM basados en machine learning e inteligencia artificial para proteger la identidad digital y los accesos no autorizados en aplicaciones web. Además se busca comparar los modelos de IAM generados con sistemas de IAM tradicionales (que no utilizan inteligencia artificial) en aspectos de seguridad y experiencia de usuario para determinar los más apropiados de implementar en aplicaciones web reales. Se pretende desarrollar e integrar los modelos basados en redes neuronales o machine learning (ML) en un sistema de gestión de identidad y acceso (IAM) de código abierto tal como Keycloak. Con esto se podría lograr un proveedor de identidad personalizado que ejecute el o los modelos de machine learning para analizar patrones de comportamiento de usuarios (por ejemplo: velocidad de tecleo, ubicación, dispositivo en uso) antes de autorizar el acceso. Esto sería posible porque Keycloak permite registrar listeners para eventos (por ejemplo: inicios de sesión, errores). Se está analizando la posibilidad de enviar datos a un servicio externo de machine learning para detectar

anomalías. En principio, se está estudiando cómo poder integrar los modelos que se vayan generando como un servicio externo y conectarlos mediante APIs para tomar decisiones en tiempo real.

2. Otro eje que se está trabajando es el diseño y ejecución de pruebas de tipo read team a los modelos y sistemas de IAM insertos en aplicaciones web. Hasta el momento se están llevando a cabo las pruebas en un entorno en Amazon Web Services, por medio de una instancia de EC2 (Amazon Elastic Compute Cloud) en la cual se despliegan el proveedor de identidad Keycloak y una aplicación web corriendo bajo el servidor nginx. Este despliegue en la nube facilita la ejecución de pruebas y posibilita también tener un proveedor de identidad y acceso disponible en caso de necesidad. También se tiene previsto el uso del servicio WSO2 Identitity Server para hacer más robustas las experimentaciones.
   Esto tiene el propósito de medir los niveles de seguridad de los mismos y obtener conclusiones.
3. Otro aspecto en lo que se está trabajando es en la investigación e implementación de un modelo de autenticación que utiliza blockchain como medio para confirmar que un usuario que intenta acceder a un servicio a través de una dirección pública pueda comprobar su identidad a través de la creación y uso de distintas direcciones públicas que actúen a modo de factores de seguridad adicionales. Esto tiene por objetivo salvar los problemas de seguridad a los que se enfrentan los usuarios de wallets en blockchains, en relación con la protección y seguridad de las claves privadas asociadas a ellas. La pérdida o robo de estas compromete de manera permanente toda la información relacionada con una determinada clave privada. La segunda arista tiene que ver con la realización de pruebas de tipo red team (como las analizadas para Web 2.0) en aplicaciones web con sistemas de autenticación basados en blockchain, más concretamente utilizando el mecanismo Sign In With Ethereum (SIWE) [7].

3. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo de la presente línea de I+D+i cuenta con la dirección del Dr. Miguel Mendez-Garabetti. Además, participan docentes investigadores de las diferentes universidades. Se cuenta con tres estudiantes de grado, y un estudiante de posgrado. Uno de los estudiantes de grado se encuentra presentando su proyecto final basado en la línea de investigación.

4. RESULTADOS ESPERADOS

Existen diferentes resultados esperados relacionados en los diferentes proyectos, los cuales confluyen al resultado esperado de la línea de I+D+i general.

El primer resultado esperado es generar un modelo de IAM autónomo, seguro y escalable.

Otro de los resultados esperados es obtener mediciones y establecer una ponderación lo más exacta posible, acerca del rendimiento y niveles de seguridad del modelo generado, con el objetivo de definir el grado de protección que ofrece a la identidad del usuario y cómo influyen en la experiencia de usuario.

Finalmente, otro resultado esperado es concluir si el uso de tecnologías Blockchain vinculadas a la Web3 mejoran los niveles de

seguridad en aplicaciones web respecto a tecnología vinculadas a las Web2.

5. BIBLIOGRFÍA

[1] R. Kaur, D. Gabrijelčič, y T. Klobučar, “Artificial intelligence for cybersecurity: Literature review and future research directions”, Information Fusion, vol. 97, p. 101804, sep. 2023, doi: 10.1016/j.inffus.2023.101804.

[2] I. A. Mohammed, “A literature review on the application of AI to Identity Access Management”, vol. 5, núm. 10, 2018.

[3] D. Cole, “Artificial intelligence and personal identity”, Synthese, vol. 88, núm. 3, pp. 399–417, sep. 1991, doi: 10.1007/BF00413555.

[4] B. Wang, D. Liu, y M. Ji, “Research on Management System of Mold Manufacturing Enterprise Based on RFID Technology”, MATEC Web of Conferences, vol. 95, p. 10002, feb. 2017, doi: 10.1051/matecconf/20179510002.

[5] Nikhil Ghadge, “Enhancing threat detection in Identity and Access Management (IAM) systems”, Int. J. Sci. Res. Arch., vol. 11, núm. 2, pp. 2050–2057, abr. 2024, doi: 10.30574/ijsra.2024.11.2.0761.

[6] S. Talluri, “AI In Identity And Access Management: Current Practices, Challenges, And Opportunities”, vol. 12, núm. 1, 2022.

[7] E. Piray, M. Méndez-Garabetti, L. Requena, R. González, G. Romero Arregin, y S. Araya, Propuesta para mejorar la protección de la identidad digital de usuarios en ambientes Web 2.0 y Web 3.0. 2024.

Uso de técnicas de Machine Learning para la Detección Temprana de Ransomware

Gonzalo Heinen, Mayra Alejandra Milano , Jorge Kamlofsky
CAETI - Centro de Altos Estudios en Tecnología Informática.
Universidad Abierta Interamericana, Facultad de Tecnología Informática

Abstract
Se explora el impacto de la convergencia entre redes de Tecnología Operacional (OT) y de Tecnologías de la Información (IT) en infraestructuras críticas, poniendo especial énfasis en la vulnerabilidad de los sistemas SCADA ante la creciente amenaza de los Ransomware. Para enfrentar este desafío, se describe un experimento que evalúa la capacidad de algoritmos de aprendizaje automático (Machine Learning) para detectar actividades maliciosas en fases tempranas de un ataque, antes de que el malware logre propagarse o cifrar datos críticos. Mediante la simulación de un entorno de red que integra componentes IT y OT de forma realista, se busca demostrar la eficacia de técnicas de análisis de comportamiento y de identificación de patrones anómalos para reforzar la seguridad de sistemas industriales estratégicos y mitigar el impacto de incidentes cibernéticos.

Palabras clave*: Infraestructuras críticas, Ciberseguridad, Redes OT, Sistemas SCADA, Ransomware, Machine Learning, Convergencia IT/OT.*

INTRODUCCIÓN

Las infraestructuras críticas desempeñan un papel fundamental en la estabilidad, la seguridad y el desarrollo de un país, puesto que abarcan sectores esenciales como la energía, el transporte, las telecomunicaciones y los servicios de defensa. En este contexto, las redes de Tecnología Operacional (OT), responsables de gestionar y controlar sistemas físicos en dichos ámbitos, han adquirido una relevancia estratégica ante la creciente necesidad de protegerlas frente a amenazas cibernéticas avanzadas [1].

La Reciente Directiva de Política de Defensa Nacional [1] enfatiza la importancia de la protección de infraestructuras estratégicas, destacando que el Sistema de Defensa Nacional debe enfocarse en la seguridad de aquellas infraestructuras cuyo funcionamiento resulta crítico para el ejercicio de la soberanía y el resguardo de la vida y libertad de los ciudadanos. En este sentido, diversas instituciones académicas y organismos de ciberseguridad han impulsado proyectos orientados a fortalecer la resiliencia de las infraestructuras OT [2], mediante el desarrollo de herramientas avanzadas de detección y respuesta ante incidentes de ciberseguridad.

En Argentina, el Ejército Argentino (EA) ha promovido el desarrollo de sistemas de comando y control (SC2) para sus brigadas, lo que ha dado lugar a una mayor interconexión y automatización de estos sistemas. Sin embargo, esta evolución también ha incrementado la superficie de ataque y la vulnerabilidad frente a amenazas cibernéticas, lo que ha motivado la creación de proyectos como la propuesta de InFoscopia [3].

Por otro lado, el Comando Conjunto de Ciberdefensa ha trabajado en la consolidación de capacidades de ciberdefensa y respuesta ante incidentes en infraestructuras críticas industriales (ICI) [4], a través de proyectos como el Programa

de Desarrollo Tecnológico-Social (PDTS)[1], que ha contado con el apoyo de diversas universidades e instituciones gubernamentales y privadas.

La relevancia de estos proyectos radica en la creciente sofisticación de los ataques a sistemas OT, que van desde la interrupción de procesos industriales hasta el secuestro de sistemas mediante ransomware y la explotación de vulnerabilidades en dispositivos Supervisory Control And Data Acquisition (SCADA).

En este contexto el objetivo del presente trabajo es el contextualizar la vulnerabilidad de los sistemas SCADA y presentar un experimento de alcance limitado que evalúe la capacidad del uso de machine learning para detectar posibles atacantes durante la fase de pre-ataque, contribuyendo así a prevenir ataques de ransomware en redes OT.

1. Convergencia entre redes IT y OT

A medida que las redes de Tecnologías de la Información (IT) y OT convergen para mejorar la eficiencia y la productividad de entornos industriales, surgen desafíos significativos en materia de ciberseguridad. Tradicionalmente, los sistemas OT funcionaban de manera aislada y estaban diseñados para operar con un énfasis casi absoluto en la disponibilidad y la seguridad física de los procesos. Por su parte, en IT se han consolidado múltiples mecanismos para proteger datos, redes y servidores frente a ataques cibernéticos. Cuando ambas dimensiones se integran, la confluencia de requisitos tan distintos provoca vulnerabilidades que ponen en riesgo la confiabilidad e incluso la seguridad de instalaciones críticas.

La amenaza del ransomware ha experimentado un crecimiento exponencial en sistemas de control industrial y otras infraestructuras críticas [6]. Este tipo de ataque, que cifra datos o inutiliza dispositivos a cambio de un rescate, ha pasado de afectar principalmente a la capa IT a propagarse cada vez más en entornos OT. La conectividad creciente de plantas industriales y la adopción de protocolos sin suficientes medidas de cifrado o autenticación ha facilitado que los atacantes encuentren nuevas vías de acceso a sistemas de supervisión y control. Tal expansión eleva el riesgo de interrupción de procesos esenciales, con consecuencias potencialmente graves en términos de producción y seguridad física.

En primer lugar, el hecho de que los dispositivos industriales (como PLC, controladores de proceso o RTU) se conecten a redes IP comunes amplía la superficie de ataque. Muchos protocolos tradicionales de campo (por ejemplo, Modbus o PROFIBUS) se diseñaron sin cifrado o autenticación adecuados, pues su objetivo original era la operación confiable y en tiempo real de procesos físicos, no la protección frente a ataques. Al migrar estos protocolos a variantes sobre TCP/IP (Modbus/TCP o PROFINET, entre otros), se heredan carencias de seguridad difíciles de remediar sin rediseñar los dispositivos o invertir en soluciones de encapsulación y monitoreo especializado.

La separación entre la red corporativa IT y la red de planta OT se ha ido diluyendo. Muchas organizaciones requieren visualizar datos de producción en tiempo real en sistemas empresariales o en la nube, lo que facilita la toma de decisiones basadas en grandes volúmenes de datos (Big Data y IIoT). Sin embargo, estos enlaces cruzados, si no se gestionan con cautela, permiten que ataques originados en la capa IT —por ejemplo, mediante correos de phishing o exploits conocidos para sistemas Windows— se propaguen hasta los equipos

[1] Decreto 703/2018. DECTO-2018-703-APN-PTE - Directiva de Política de Defensa Nacional.

[2] Proyectos de desarrollo tecnológico y/o de impacto social.

de control industrial. A la luz de la escalada de casos de ransomware mencionada, esta interconexión se convierte en un factor crítico que agrava la exposición de los activos OT [5].

2. FASES DE ATAQUE DE UN RANSOMWARE

Las fases de un ataque de ransomware comienzan antes incluso de la ejecución del malware en los sistemas de la víctima. En esta etapa de pre-ataque, los ciberdelincuentes llevan a cabo labores de reconocimiento y obtención de información, conocidas como reconocimiento (reconnaissance) en la Cyber Kill Chain. Durante esta fase preliminar, los atacantes identifican posibles vulnerabilidades en los sistemas de la organización, recopilan credenciales filtradas en la dark web, estudian qué servicios están expuestos a Internet e investigan las medidas de seguridad implementadas. Asimismo, pueden preparar herramientas de acceso inicial como troyanos o correos de phishing que servirán para la fase de entrega (delivery), cuidando la furtividad de sus métodos y adaptando sus tácticas según el perfil de la víctima.

Una vez logrado el acceso, los atacantes se enfocan en escalar privilegios, moverse lateralmente por la red y establecer persistencia en los sistemas comprometidos. Para ello, pueden emplear técnicas como exploits conocidos para servidores o el uso de credenciales robadas. Completado este movimiento interno, se pasa a la fase de impacto o ataque directo, en la que el ransomware cifra ficheros críticos y, a menudo, exfiltra datos sensibles para ejercer presión (doble extorsión). Finalmente, la víctima recibe la nota de rescate en la que se detalla el pago exigido para restaurar sus datos o evitar su publicación [7].

3. MACHINE LEARNING PARA LA DETECCIÓN DE UN RANSOMWARE

La implementación de técnicas de Machine Learning para la detección de ransomware podría representar un paso relevante en la búsqueda de soluciones más adaptables y robustas para proteger infraestructuras informáticas. Se cree que, al emplear algoritmos como Random Forest, Naïve Bayes o Support Vector Machine, se lograría analizar patrones de comportamiento en archivos o procesos que, tradicionalmente, pasan inadvertidos en enfoques basados únicamente en firmas o reglas fijas. De concretarse, dichos modelos tendrían la capacidad de aprender continuamente de nuevos ataques, ampliando su eficacia frente a variantes más recientes de malware y, en particular, de ransomware.

Sin embargo, la implementación de esta solución presenta problemas. Uno de los principales es el costo/beneficio de la implementación de un hardware lo suficientemente potente para emplear el machine learning de forma óptima e inmediata [8].

4. LÍNEAS DE INVESTIGACIÓN Y EXPERIMENTO EN REDES IT/OT

En el marco del proyecto de Ciberdefensa en redes OT del CAETI, el objetivo es estructurar un experimento diseñando una infraestructura de red interna que emula de manera realista las condiciones y registros característicos de un entorno de oficina, con el fin de estudiar y anticipar amenazas de tipo ransomware. Para ello, se han dispuesto dos segmentos de red (IT y OT) tal como se aprecia en las topologías, donde cada

segmento aloja diversos equipos y servicios comunes en ambientes corporativos (por ejemplo, servidores, puestos de trabajo y dispositivos de automatización). Este montaje controlado posibilita la captura y el análisis de eventos de manera más fidedigna, reflejando con mayor precisión la actividad esperable en escenarios de producción.

Bajo este contexto, la intención es aplicar en la siguiente etapa del experimento técnicas de inteligencia artificial, concretamente de aprendizaje automático (Machine Learning), para revisar sistemáticamente los registros y flujos de datos generados en la red. De esta manera, se espera identificar patrones de comportamiento anómalo asociados con un potencial ataque de ransomware, particularmente desde la fase de pre-ataque. En última instancia, el objetivo consiste en validar la capacidad de un modelo automatizado para reconocer señales tempranas de infección antes de que el malware logre propagarse y comprometer los sistemas críticos, tanto en redes IT/OT separadas como unificadas.

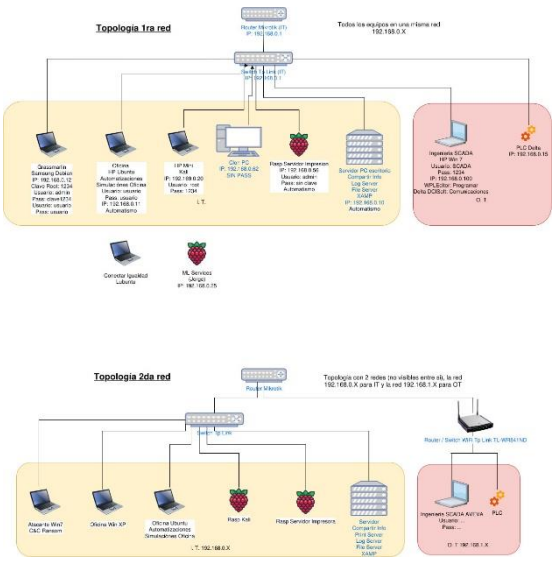


Figura 1. “TOPOLOGÍA DE REDES 1 Y 2”

Se observan las topologías de redes empleadas para dos casos de estudios distintos. En el primer caso la red OT se encuentra en convergencia con la red IT, mientras que en el segundo se encuentran por separado.

5. FORMACIÓN DE RECURSOS HUMANOS

El presente proyecto se encuentra dirigido por el Mg. Lic. Jorge Kamlofsky quien está cursando un Doctorado. Los resultados colaborarán con el desarrollo de su Tesis.
Para el desarrollo de las actividades está prevista la participación de dos estudiantes de grado de la Universidad Abierta Interamericana de la carrera de grado de Ingeniería en Sistemas de información.
Por otro lado, el presente proyecto se enmarca como una de las líneas de trabajo que viene desarrollando el Laboratorio CAETI de la Universidad, en donde alumnos de grado y posgrado realizan sus trabajos finales de carrera. Por lo tanto, está prevista la incorporación de dos alumnos de grado y posgrado, quienes profundizarán sus saberes y realizarán los aportes correspondientes.

6. BIBLIOGRAFÍA

[1]: Giorgio Valenziano Santangelo y Vincenzo Giuseppe Colacino, “Analysis, prevention and detection of ransomware attacks on Industrial Control Systems”, Conference: 2021 IEEE 20th International Symposium on Network Computing and Applications (NCA)

[2]: Jorge Kamlofsky, Hugo Colombo, Claudio Milioy Pedro Hecht. “Ciberdefensa en Sistemas Operacionales”. Workshop de Investigadores en Ciencias de la Computación WICC 2024.

[3]: Liporace, Julio César, Buscaglia, Adrián, Croci, Pablo, Díaz Pais, Nicolás, Fernández, Darío, Ferreyra, Verónica, Gallardo Urbini, Ignacio Martín, Quiroga, Elvira, Vera Batista, Fernando, Cicerchia, César D. “Metodología para el análisis de incidentes de ciberseguridad o ciberataques durante las acciones de ciberdefensa de las infraestructuras críticas de la defensa nacional –infoscopia–“. Tecnología Workshop de Investigadores en Ciencias de la Computación WICC 2019

[4]: Kamlofsky, Jorge, Gonzalez, Gerardo, Trigo, Santiago “Infraestructuras Críticas Industriales ICI”. Workshop de Investigadores en Ciencias de la Computación WICC 2021.

[5]: Georgios Michail Makrakis, Constantinos Kolias, Georgios Kambourakis, Craig Rieger, Jacob Benjamin," Vulnerabilities and Attacks Against Industrial Control Systems and Critical Infrastructures", Cornell University 10 Sep 2021.

[6]: Waterfall. "Industrial Cybersecurity" 2024.

[7]: Ramón José Paniagua Soza. Anatomía del ransomware. Junio 2022 Recuperado el 21 de febrero del 2025.

· TIAE ·

Tecnología Informática Aplicada en Educación

Alfadatizando: Enseñanza en nivel secundario de Pensamiento Computacional para Humanidades Digitales

Alejandra B. Lliteras [1,2], Alejandro Artopoulos [2,3,4], Alejandro Fernández [1,2]

[1] UNLP, Facultad de Informática, Centro LIFIA, [2] CICPBA,

[3] UDESA, Escuela de Educación, [4] UBA

alejandra.lliteras@lifia.info.unlp.edu.ar, alepoulos@udesa.edu.ar, alejandro.fernandez@lifia.info.unlp.edu.ar

RESUMEN

El trabajo presenta una línea de investigación en curso que se enmarca en un tema doctoral con el fin de desarrollar pensamiento computacional a través de la visualización de datos para el área de humanidades digitales en el nivel secundario.

Para la adopción de las humanidades digitales en el nivel secundario, es necesario establecer estrategias de enseñanza para las diferentes disciplinas que involucren ciertos métodos computacionales, permitiendo desarrollar el pensamiento computacional de manera disciplinar, con el fin de poder reflexionar sobre cómo utilizar la tecnología digital y sobre lo que se hace a través de ella, además de utilizarse para pensar en cómo proponer una solución. Esto no sólo permite desarrollar nuevas formas de conocimiento y prácticas en las humanidades, sino que también contribuye a una ciudadanía más crítica y reflexiva.

CONTEXTO

Esta línea de investigación continua con la propuesta presentada en este Workshop en el año 2024 [Lliteras, 2024] desde la importancia de abordar la construcción de ciudadanía digital desde diferentes disciplinas y áreas curriculares (continuación de [Lliteras et al., 2023]), en particular, haciendo foco en las Humanidades Digitales. La temática, se enmarca en un proyecto de doctorado en curso [Lliteras, 2020].

1- INTRODUCCIÓN

El campo multidisciplinar de las Humanidades Digitales (HD) se desarrolló como una práctica de investigación que investiga y aplica métodos digitales en la producción de conocimiento en humanidades y ciencias sociales, para luego aplicarlos también en la enseñanza de dichas disciplinas en el nivel superior, y en algunos casos en el nivel secundario.

Berry y Fagerjord desde una perspectiva disciplinar, mencionan que las humanidades digitales requieren de un enfoque crítico de pensamiento, el cual puede llevarse adelante con el pensamiento computacional [Berry & Fagerjord, 2017].

En las humanidades digitales se utilizan a menudo métodos cualitativos de las humanidades, pero también métodos cuantitativos de las ciencias sociales y la informática, métodos computacionales en general como el *storytelling*, la lectura distante [Moretti, 2013], el análisis de redes (grafos), la inteligencia artificial generativa y la visualización de datos [Mills, 2020]. Cuando se trabaja con humanidades digitales, es posible emplear el pensamiento computacional para reflexionar sobre cómo utilizar la tecnología digital y sobre lo que se hace a través de ella, además de utilizarse para pensar en cómo proponer una solución. Este tipo de pensamiento en las humanidades digitales transforma la argumentación y la forma de pensar el conocimiento. En su trabajo, Berry y Fagerjord reflexionan sobre la necesidad de aprender sobre humanidades digitales y cómo el futuro de su enseñanza debe considerar la importancia de adoptar a la tecnología de la información desde la perspectiva de la disciplina y entender que, con este tipo de

pensamiento, lo que ocurre computacionalmente, es que surge la posibilidad de que los ciudadanos puedan pensar no sólo en usar, sino también en modificar o crear cosas diferentes a las existentes. Los autores también mencionan que el desarrollo de la sensibilidad disciplinar no sólo permite desarrollar nuevas formas de conocimiento y prácticas en las humanidades, sino que también contribuye a una ciudadanía más crítica y reflexiva que puede utilizar estos aspectos para contribuir a la sociedad, teniendo en cuenta que las humanidades digitales crean nuevos puestos de trabajo, habilidades y conocimientos específicos para una nueva economía [Berry & Fagerjord, 2017], redundando en la formación de los estudiantes como ciudadanos digitales [Ribble, 2008].

Esta nueva forma de hacer humanidades requiere de nuevas estrategias de enseñanza, y si bien existen diversos estudios que abordan la enseñanza de las humanidades digitales en el nivel terciario y universitario, muy pocos lo hacen para nivel primario o secundario [Pérez García, 2021]. Por otro lado, para avanzar en el debate de la pedagogía de las humanidades digitales es necesario que más estudios se focalicen es este nivel [Carnes and Smith, 2024].

En base a lo anterior, en esta línea de investigación se hace foco en el desarrollo del pensamiento computacional a través de la visualización de datos para el área de humanidades digitales en el nivel secundario

2- LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

La línea de investigación presentada apunta al estudio de estrategias de enseñanza de pensamiento computacional a través de la visualización de datos en disciplinas de humanidades para fortalecer la formación de ciudadanos digitales considerando tanto la alfabetización digital como la alfabetización en datos analizando y generando plataformas y tecnologías que permitan la innovación en el aula para favorecer el proceso de enseñanza aprendizaje en el marco del nivel secundario.

3- RESULTADOS OBTENIDOS

Como parte de los resultados obtenidos se cuenta con Alfadatizando[1], una plataforma de acceso abierto para la generación de actividades educativas con datos para el área de humanidades digitales la cual se usó como base para la nueva versión desarrollada (Alfadatizando 2.0) que en la actualidad se encuentra en etapa de testing. Alfadatizando 2.0 permite fuentes de datos variadas, ya sean estáticas (tablas, grafos, archivo pdf) y dinámicas (datos de la red social Reddit[2]). Por otro lado, es posible trabajar con diferentes tipos de gráficos como, por ejemplo, nubes de palabras y grafos.

Por otro lado, se presentó el informe final y se defendió una tesina de grado en la que se propuso una arquitectura para la recolección de datos a través de sensores y su posterior visualización y análisis desde un servidor.

Como parte del trabajo realizado con esta línea de investigación se realizaron diversas publicaciones en revistas, congresos y jornadas y se dictó un taller que presentaba el abordaje de la enseñanza de la lectura distante usando herramientas abiertas y tomando como base grafos de obras de la literatura española. Los asistentes fueron docentes secundarios del Plan Ceibal (Uruguay) y docentes universitarios/investigadores (de diferentes países). Los resultados de este taller están pendientes de publicación.

[1] https://link.springer.com/book/9789819636976?srsltid=AfmBOorbGxPaPCgpdgrqWJrsz8arijvaRg-w3qhvNw25vYBbI3zMXu3zhttp://www.alfadatizandonos.okd.lifia.info.unlp.edu.ar/

[2] https://www.reddit.com/

Se realizó la edición de un proceeding en Lecture Notes In Educational Technology de Springer para articulos presentados en el Congreso LACLO 2024.

Respecto a las publicaciones realizadas, las mismas se listan a continuación:

Artículos en revistas [Artopoulos & Lliteras, 2024a], [Artopoulos & Lliteras, 2024b]. Artículos en congresos [Lliteras et al., 2024], [Lozano et al., 2024]. Artículos en workshops [Lliteras, 2024]. Articulos en jornadas [Artopoulos & Lliteras, 2024c], [Fernández et al., 2024].

4- RESULTADOS ESPERADOS

Se espera en el transcurso de los próximos meses cerrar la tesina de grado que considera aspectos propios de la ingeniería de software en relación con la plataforma Alfadatizando 2.0. De manera simultánea, se iniciará una tesina de grado que continuará aportando a la plataforma (Alfadatizando 3.0) sumando otros métodos computacionales.

En el corriente año se inició el proceso de una tesis de maestría cuyo foco es el desarrollo de pensamiento computacional en la enseñanza de la lengua inglesa.

Se espera la publicación de artículos que fueron aceptados durante el 2024 pero que aún se encuentran en proceso de edición:

Artículo en Revista de Lenguas Vivas[3], llamado “Aproximaciones a la alfabetización crítica en IA en la enseñanza de la literatura en el nivel secundario”.

Artículos en Congreso LACLO 2024[4], por un lado “Alfadatizando 2.0 applied to data visualization at high school level and for digital humanities. Empowering digital citizens”, por otro, “Teaching digital humanities at high school level. A literature and special cases review”

Ponencia realizada en XII Jornada Socio UNLP[5] “La construcción social de la ciudadanía digital. Movimientos sociales, activismo experto y cambio curricular”.

En base al taller dictado durante el 2024 se espera publicar los resultados y proponer un nuevo taller orientado a docentes. También se espera realizar entrevistas con informantes clave del sistema educativo en relación con prácticas docentes en humanidades y ciencias sociales.

Finalmente, se espera continuar en el avance de la tesis doctoral en curso.

5- FORMACIÓN DE RECURSOS HUMANOS

La actualidad del tema, así como la diversidad de líneas de investigación previamente presentadas, muestra una línea de investigación concreta y relevante para la formación de recursos humanos, con los cuales se espera trabajar tanto para tesinas de grado, maestría, trabajos de títulos intermedio como con pasantes.

Tesina de grado aprobada: 1 estudiante de grado. Tesina de grado en curso: 2 estudiantes. Tesina de grado en curso: 1 estudiante. Tesis de maestría en curso: 1 estudiante. Tesis doctoral

6- BIBLIOGRAFÍA

[Artopoulos & Lliteras, 2024a] Artopoulos A., Lliteras A. La emergencia de la alfabetización crítica en IA: la reconstrucción social de la ciudadanía digital bajo acecho digital. Revista Dialogo Educacional, v.24, n.83, pp. 1283-1305. Editora Universitaria Champagnat, Curitiva, Brasil. 2024

[3] https://ieslvf-caba.infd.edu.ar/sitio/revista-lenguas-vivas/

[4] https://laclo2024.org/

[5] http://jornadassociologia.fahce.unlp.edu.ar/

[Artopoulos & Lliteras, 2024b] Artopoulos, A., Lliteras, A.: Alfabetización crítica en IA: Recursos educativos para una pedagogía de la descajanegrización. Trayectorias Universitarias, 10(19), p. 168. UNLP, La Plata. 2024

[Artopoulos & Lliteras, 2024c] Artopoulos A., Lliteras A.: Enseñanza de las Humanidades Digitales en escuelas secundarias. In: Humanos que piensan con máquinas. Sexto Congreso Internacional de la Asociación Argentina de Humanidades Digitales. Asociación Argentina de Humanidades Digitales (AAHD), Instituto de Investigaciones Geohistóricas (Consejo Nacional de Investigaciones Científicas y Técnicas - Universidad Nacional del Nordeste). 2024

[Berry & Fagerjord, 2017] Berry, D. M., & Fagerjord, A. Digital humanities: Knowledge and critique in a digital age. John Wiley & Sons. 2017

[Carnes & Smith, 2024] Carnes, G., & Smith, M. K. Bridging the Divide: Improving Digital Humanities Peda-gogy by Networking Higher Education and Secondary Education Faculty in St. Louis. IDEAH, 4(2). 2024

[Fernández et al., 2024] Fernández G.P, Martínez López P., Artopoulos A. , Lliteras A. Reflexiones sobre la naturaleza de la Computación. Jornadas Argentinas de Didáctica de las Ciencias de la Computación .Facultad de Ciencias Exactas Físico-Químicas y Naturales de la Universidad Nacional de Río Cuarto. 2024

[Lliteras et al., 2022] Lliteras Alejandra, Artopoulos Alejandro, Fernandez Alejandro., & Huarte Jimena. AlfaDatizando: a Data Visualization Platform to work Computational Thinking in Digital Humanities. In 2022 XVII Latin American Conference on Learning Technologies (LACLO) (pp. 1-6). IEEE. 2022

[Lliteras et al., 2023] Lliteras A.B., Artopoulos A, Fernández A. & Huarte J. Alfadatizando: innovación en el proceso de enseñanza aprendizaje en escuelas secundarias usando visualización de datos para desarrollar pensamiento computacional en materias de Humanidades y Ciencias Sociales. Libro de Actas WICC 2023- UNNOBA: XXV Workshop de Investigadores en Ciencias de la Computación. (pp. 895-897). 2023

[Lliteras et al., 2024] Lliteras A., Lozano Arce J., Rodriguez A.: Recolección, envío, almacenamiento y visualización de datos considerando programación basada en bloques de placas MicroBit, ESP, sensores y Wifi. Libro de Actas - 30° Congreso Argentino de Ciencias de la Computación - CACIC 2024. pp.1592-1596. Facultad de Informática (UNLP), La Plata. 2024

[Lliteras, 2020] Lliteras A., “Aprendo Sociales con Datos. Con el fin de construir conocimiento, en Ciencias Sociales, a partir del análisis de conjuntos de datos usando modelización, visualización y razonamiento algorítmico para escuelas secundarias de Argentina.”. Propuesta doctoral. UNLP. Facultad de Informática. 2020

[Lliteras, 2024] Lliteras A.: Alfadatizando como parte de la Ciudadanía Digital. Libro de Actas XXVI Workshop de Investigadores en Ciencias de la Computación. pp. 823-826. Universitaria de la Patagonia -EDUPA, Comodoro Rivadavia. 2024

[Lozano et al., 2024] Lozano Arce J., Lliteras A., Rodriguez A.: IRID: Infraestructura para la recolección inalámbrica de datos provistos por sensores en el marco de robótica educativa en nivel secundario. Libro de Actas - 30° Congreso Argentino de Ciencias de la Computación - CACIC 2024. pp.1592-1596. Facultad de Informática (UNLP), La Plata. 2024

[Mills, 2020] Mills, T. E. Building a Pedagogical Relationship between Philosophy and Digital Humanities through a Creative Arts Paradigm. Teaching Philosophy, 43(4), 403-429 (2020).

[Moretti, 2013] Moretti, F. Distant reading. Verso Books. 2013

[Pérez García, 2021] Pérez García, C. Humanidades digitales y educación literaria: oportunidades y retos. 2021

[Ribble, 2008] Ribble, M. Passport to digital citizenship. Learning & leading with technology, 36(4), 14-17. 2008

Análisis de la movilidad estudiantil. Formulación de casos de estudio basado en datos de becas Progresar y Belgrano

Gustavo Illescas[1], Elías Todorovich[1], Thomas Ojeda[2], Gabriel Montaruli

[1]Instituto de Investigación en Tecnología Informática Avanzada (INTIA-Centro Asociado CIC, Facultad de Ciencias Exactas, Universidad Nacional del Centro de la Provincia de Buenos Aires (UNCPBA).
[2]Becario EVC-CIN, Facultad de Ciencias Exactas, UNCPBA.
[3]Centro de Información y Estadística Universitaria, Rectorado UNCPBA.

{illescas, etodorov}@exa.unicen.edu.ar, tojeda@alumnos.exa.unicen.edu.ar, gmontar@rec.unicen.edu.ar

RESUMEN

Este trabajo propone la formulación de casos de estudio, basados en datos de las becas Progresar y Manuel Belgrano, con el fin de generar un tablero de control con indicadores clave para la toma de decisiones. Para esto se ha comenzado con el desarrollo de una herramienta web que permita, con los datos recolectados, analizar la movilidad estudiantil y buscar patrones de comportamiento. Se espera que este análisis pueda ser escalado a nivel nacional, lo que facilitará la difusión de los resultados y herramientas desarrolladas.
El tema se encuentra enmarcado dentro del proyecto de incentivos 03/C314 titulado "Gestión Informática del Conocimiento como soporte para la toma de decisiones Organizacionales" desarrollado en el Instituto de Tecnología Informática Avanzada (INTIA) de la Facultad de Ciencias Exactas (EXA) UNCPBA y ha sido presentado y aprobado como beca EVC-CIN, en el año 2024.

Palabras clave: Movilidad Estudiantil, Educación Superior, Analítica de Datos, Indicadores.

CONTEXTO

Es de interés para el Centro de Información y Estadística Universitaria (CIEU) de la Universidad Nacional del Centro de la Provincia de Buenos Aires (UNCPBA) obtener una herramienta que permita calcular ciertos indicadores de comportamiento de la población estudiantil y plasmarlos gráficamente con el objetivo de asistir a la toma de decisiones, de forma similar a un tablero de control (Quesado, et. al. 2012, Illescas, et. al. 2018).

Los indicadores que usualmente se usan para medir las trayectorias académicas son el avance académico, el rendimiento académico, la eficiencia terminal, el retraso y la deserción académica, la regularidad y el fracaso (Sneyers, et. al. 2017). Durante los últimos 3 años, la síntesis de información universitaria (Alpa, 2022) ha incluido la "Tasa de cambio entre ofertas académicas", que representa el porcentaje de estudiantes que se han inscrito a carreras de grado de instituciones universitarias de gestión estatal, y que optan por otra oferta académica uno o dos años después de su ingreso. Sin embargo, estos indicadores no se encuentran disponibles en nuestra institución. Se tiene además un interés en conocer su correlación con la adjudicación de becas nacionales, actualmente Becas Manuel Belgrano y Becas Progresar. El valor de este indicador puede representar el comportamiento de los estudiantes en situaciones que anteriormente se clasificaban como deserciones, y que se sospecha se encuentran en movimiento entre ofertas con el objetivo de mantener los requisitos de su correspondiente Beca.

En este contexto, se desarrolló durante la asignatura optativa Científicos de Datos (EXA) utilizando el *framework* Shiny (Web Shiny, 2023), una aplicación web con formato de tablero de control que permite la visualización de la tasa de cambio de los estudiantes de las distintas ofertas académicas de la UNCPBA, utilizando como fuente de datos archivos en formato de planilla de cálculo provistos por el CIEU.

1. INTRODUCCIÓN

Las entidades e instituciones públicas llevan a cabo un amplio registro de datos sobre el comportamiento de su población. El objetivo principal de este registro es contribuir a la mejora de las políticas públicas. Dentro de este registro, se encuentran estadísticas relacionadas con el sistema educativo las cuales ofrecen una visión general de la situación actual del sistema educativo. Además, permiten comparar mediciones anteriores para establecer una cronología de su progreso que facilita la formulación de planes y fomenta el desarrollo del sistema educativo (Canales-Sánchez, 2022).

En nuestro país, la Secretaría de Políticas Universitarias Nacional presenta anualmente la Síntesis de Información Universitaria (Alpa, 2022), una publicación elaborada por el Departamento de Información Universitaria, perteneciente a la Dirección Nacional de Presupuesto e Información Universitaria. Esta publicación contiene los últimos datos disponibles del sistema universitario argentino, tales como indicadores sobre las instituciones universitarias, la población estudiantil de pregrado, grado y posgrado y las ofertas académicas para la totalidad del sistema universitario. Los datos utilizados son proporcionados por las instituciones a través de los diversos relevamientos que se realizan a lo largo del año.

La elaboración de estos informes requiere emplear conceptos de la Ciencia de Datos, la cual abarca los campos de la matemática y estadística, programación, inteligencia artificial y *machine learning* con la experticia en dominios específicos. Esto con el objetivo de desentrañar características útiles escondidas en los datos, las cuales pueden ser usadas para guiar la toma de decisiones y el planeamiento estratégico (Noguera, 2007).

Los datos generados principalmente por motivos administrativos en instituciones educativas acerca de la trayectoria académica de los estudiantes universitarios son valiosos para el cálculo y análisis de diversos indicadores. Estos son útiles para asistir a la toma de decisiones y la mejora de los procesos de formación y administración académica (Canales-Sánchez, 2022).

Tomando como prototipo la aplicación anteriormente descrita, se plantea una extensión de la misma para que permita visualizar más indicadores y que los mismos usen como fuente de datos archivos que se puedan incluir dinámicamente, conforme estos se hagan disponibles. De manera complementaria se plantea la utilidad de implementar algún sistema de *login* que facilite el acceso a la información de manera controlada. Es otro requisito construir esta aplicación utilizando tecnologías establecidas actualmente en el campo del desarrollo web tanto del lado servidor (Haro, 2019), como del lado cliente (presentación en el navegador web) (Jadhav et. al. 2015) para facilitar su mantenimiento y extensión.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Con las motivaciones descritas, el presente proyecto se formaliza en un plan de tesis de Ingeniería en Sistemas y tiene como objetivo diseñar, implementar y poner en funcionamiento una aplicación web cuyo principal propósito sea facilitar la visualización de estadísticas de los estudiantes. Con tal objetivo, la plataforma debe permitir cargar y realizar un procesamiento automático de archivos de datos, a partir de los cuales se pueda generar información derivada para confeccionar visualizaciones. Para regular la transmisión de los gráficos que se puedan

considerar de interés, se incluirá un método de autenticación de usuarios y un método de autorización que habilite o restrinja el acceso de cada usuario a un subconjunto específico de estadísticas generadas. Este manejo de usuarios estará regulado por uno o varios usuarios con permisos especiales.

Respecto a las tecnologías y metodologías utilizadas para el desarrollo del prototipo se compone de 4 capas bien definidas que se comunican a través de interfaces del tipo REST (*representational state transfer*). La capa encargada de la presentación fue desarrollada utilizando el *framework* de aplicaciones *front-end* SPA Angular versión 2+, el cual provee

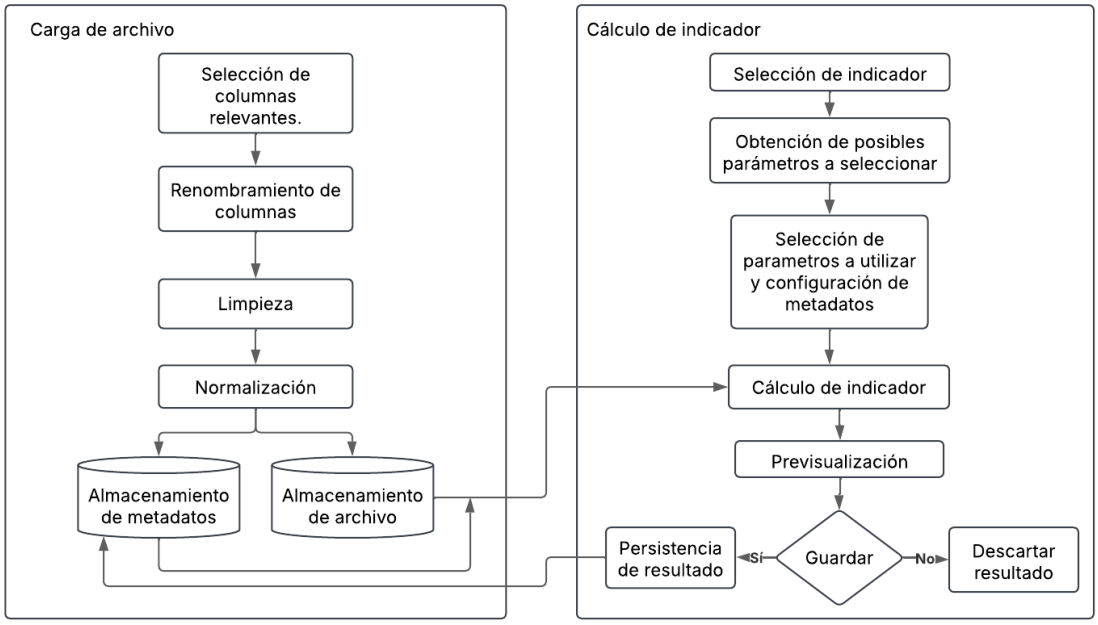


Figura 1. Diagrama de actividades del prototipo.

Inicialmente se incluirá el cálculo de los siguientes indicadores:

- Cambios entre ofertas académicas.
- Cambios entre ofertas académicas considerando adjudicación de becas nacionales.
- Carreras o unidades a las que pertenecen los estudiantes con becas adjudicadas.
- Destino de los estudiantes que cambian de oferta.
- Inscripciones y reinscripciones anuales a las distintas ofertas o unidades académicas.

Metodología

Siguiendo bajo la misma línea de trabajo que el proyecto acreditado (Illescas G., et. al 2022), se utilizará la misma metodología de desarrollo *Design Thinking* (Brown T., 2009), siendo un método de trabajo en equipo que propicia la colaboración y la entrega frecuente de resultados a través de varias iteraciones.

toda la infraestructura requerida para ofrecer una experiencia de usuario atractiva, funcional y accesible, utilizando, a lo largo de todo el proceso, buenas prácticas en la escritura del código lo que garantiza la escalabilidad. La capa encargada del control fue desarrollada utilizando el motor NodeJS valiéndose de la librería de API ExpressJS, lo que provee una gran cuota de libertad y flexibilidad para definir el esquema de comunicación. Por su parte, la capa encargada de la transformación de datos fue desarrollada utilizando un servidor Python facilitado por la librería Flask, la cual define su propia API que es consumida por el controlador. El servidor Python se vale de una librería bien conocida en el mundo del análisis de datos llamada Pandas, la cual utiliza para realizar las operaciones de transformación de datos. Por último, la capa de persistencia de datos se vale del motor MongoDB NoSQL + Mongoose, cuyo esquema flexible nos permite atender cualquier tipo de requerimiento que

pudiera surgir en materia de almacenamiento de metadatos, usuarios o el estado de la aplicación.

3. RESULTADOS OBTENIDOS/ESPERADOS

Al momento del envío de esta publicación se cuenta con un primer prototipo de la aplicación web planificada que se encuentra en fase de testeo por parte del personal del CIEU.

En la figura 1 se observa el diagrama de actividades para la carga de archivos al prototipo y la creación de los indicadores para su posterior visualización.

La estructura del equipo de trabajo se muestra en la siguiente tabla:

Apellido y nombre	Título	Cargo	Funciones
Illescas, Gustavo	Dr.	Prof. UNCPBA	Director
Todorovich, Elías	Dr.	Prof. UNCPBA	Co-Dir.
Ojeda, Thomas	Pre-grado	Alumno	Becario
Aciti, Claudio	Mg.	Prof. UNCPBA	Integrante
Rodríguez, Guillermo	Dr.	Prof. UNCPBA	Integrante
Silvestrini, Pía	Med. Vet.	Prof. UNCPBA	Integrante

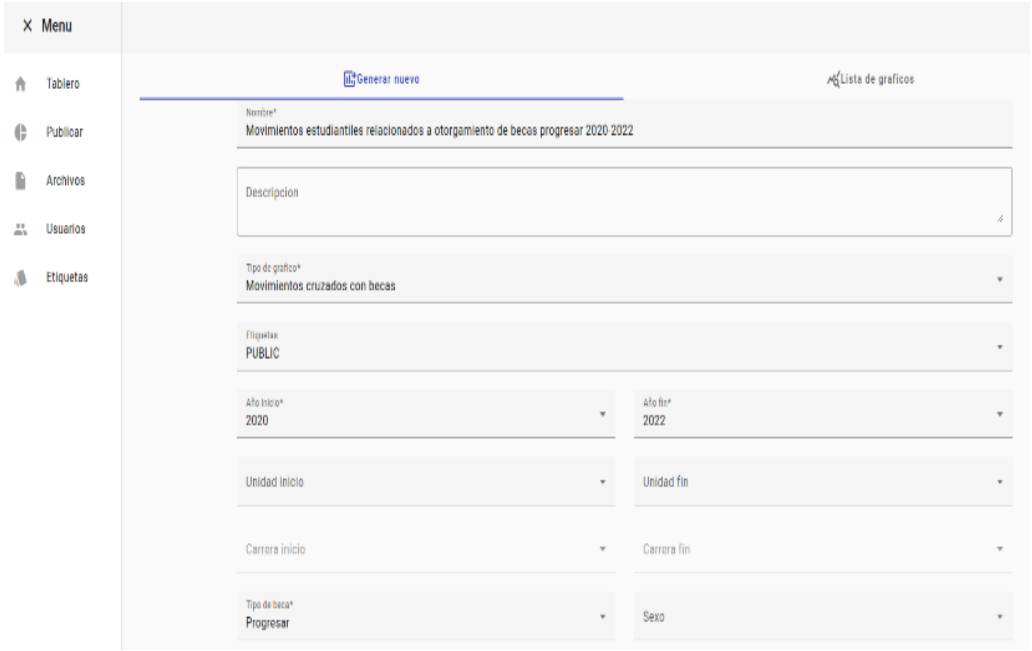


Figura 2. Configuración de un indicador.

La figura 2 muestra la configuración de un indicador a ser visualizado mediante un gráfico.

Como se observa en la figura 3, luego de procesado el archivo se puede observar el gráfico que representa el indicador calculado.

Seguidamente a la devolución de los usuarios potenciales se procederá a las modificaciones que resulten convenientes como así también a finalizar con la fase de escritura del trabajo de tesis y su posterior defensa por parte del becario.

4. FORMACIÓN DE RECURSOS HUMANOS

Becario y tesis de grado

- *Sistematización de la movilidad estudiantil como atenuante de la deserción* (obtenida en 2024). Dirección: Illescas G., Todorovich E. Becario: Ojeda Thomas (EVC-CIN 2024).

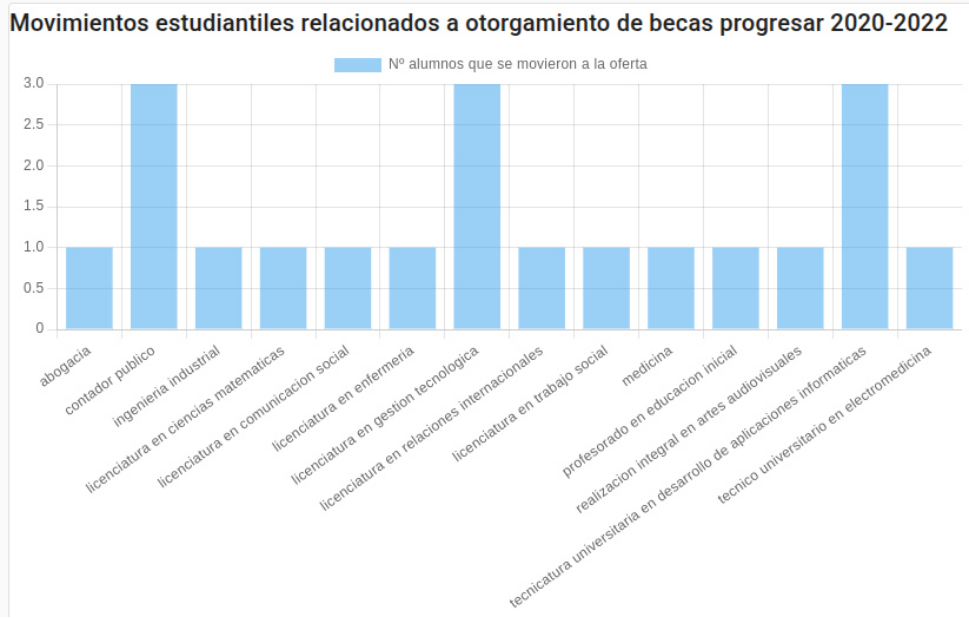


Figura 3. Visualización del indicador en el prototipo.

5. BIBLIOGRAFÍA

- Quesado P., Aibar-Guzmán B., Lima Rodrigues L. (2012). El Cuadro de Mando Integral como Herramienta de Gestión Estratégica del Conocimiento. Revista Perspectivas em Gestão & Conhecimento. 2. 70-102.
- Illescas G., Santiago M., Martínez, M. (2018). Procedimientos y herramientas para la mejora de indicadores académicos en facultades de ingeniería. XX Workshop de Investigadores en Ciencias de la Computación, Universidad Nacional del Nordeste.
- Sneyers E., De Witte K. (2017). The interaction between dropout, graduation rates and quality ratings in universities. Journal of the Operational Research Society Internet]. doi.org/10.1057/jors.2016.15
- Alpa, O. (2022). Síntesis de Información Universitaria 2021-2022 - Publicación de la Secretaría de Políticas Universitarias Nacional Síntesis 2021-2022. (https://www.argentina.gob.ar/sites/default/files/sintesis_2021-2022_sistema_universitario_argentino_1.pdf visitado en mayo 2023)
- Web Shiny (2023) Sitio web del framework de ciencia de datos (www.posit.co visitado en noviembre 2023)
- Noguera P. (2007). How Listening to Students Can Help Schools to Improve. Theory into Practice, 46(3), 205–211. doi:10.1080/14241270701402165
- Haro E., Guarda T., Peñaherrera A., Quiña, G. (2019). Desarrollo backend para aplicaciones web, servicios web restful: Node. js vs spring boot. Revista Ibérica de Sistemas e Tecnologías de Informação, (E17), 309-321.
- Jadhav M., Sawant B., Deshmukh A. (2015). Single page application using angularjs. International Journal of Computer Science and Information Technologies, 6(3), 2876-2879.
- What is Data Science? - Artículo en línea de IBM (visitado en junio 2024) https://www.ibm.com/think/topics/data-science
- Canales-Sánchez D., Godínez T., Moreno-Salinas G., García-Minjares M., Sánchez-Mendiola M. (2022). Academic trajectories analysis with a life-course approach: A case study in medical students. Cogent Education (2022), 9: 2018118
- Illescas Gustavo, Todorovich Elías, Aciti Claudio, Rodriguez Guillermo, Silvestrini Pía (2022). Aplicación de Analítica de Datos en espacios de Educación Superior. XXIV Workshop de Investigadores en Ciencias de la Computación Universidad Champagnat. Mendoza, Argentina.
- Brown, Tim (2009). Change by Design, How Design Thinking Transforms Organizations and Inspires Innovation. Ed. Harper Collins New York ISBN 978-0-06-193774-3

APLICACIÓN DE LABORATORIOS VIRTUALES EN EDUCACIÓN DE CIENCIAS EXACTAS

Godoy Pablo Daniel[1,2,3,4], Marianetti Osvaldo Lucio[1,4], Fontana Daniel[1,4], Manganelli Silvina[2]
[1]Universidad Nacional de Cuyo, Facultad de Ingeniería
[2]Universidad Nacional de Cuyo, Facultad de Ciencias Exactas y Naturales
[3]Universidad Nacional de Cuyo, ITIC
[4]Universidad de Mendoza
pablo.godoy@ingenieria.uncuyo.edu.ar

RESUMEN

El presente artículo describe un proyecto aprobado y financiado por la Universidad Nacional de Cuyo, cuyo plazo de ejecución es entre 2022 y 2024.

El proyecto tiene como objetivos el desarrollo de laboratorios virtuales para ser utilizados en tareas docentes y realizar aportes en el campo de los laboratorios virtuales o su aplicación en tareas de enseñanza .

El proyecto se encuentra en su etapa final de ejecución, habiéndose desarrollo tres laboratorios virtuales, uno de efecto Doppler y uno de ondas mecánicas para la enseñanza de Física, y uno de ventanas deslizantes para la enseñanza de redes de computadoras.

Área: TIAE (Tecnología Informática Aplicada en Educación)

CONTEXTO

El presente trabajo describe un proyecto de investigación y desarrollo financiado por la Universidad Nacional de Cuyo a través de los proyectos "SIIP B008: Desarrollo y aplicación de laboratorios virtuales destinados a educación de grado y pregrado" y "SIIP B036-T1: La seguridad de los dispositivos utilizados en IoT mediante la utilización de Blockchain, NFT y/o VPN". El plazo de ejecución inicial es entre 2022 y 2024, habiéndose extendido a 2025.

El proyecto surge como una evolución de dos proyectos de investigación y la integración de un trabajo de tesis doctoral. Dichos proyectos se denominan "Implementación de laboratorios remotos basados en cloud computing" y "Laboratorio remoto y nómada de arquitectura de computadoras destinado a enseñanza e investigación".

Estos laboratorios remotos y nómadas fueron utilizados para que estudiantes de distintas carreras realicen sus trabajos prácticos. Los laboratorios virtuales han sido empleados en: Licenciatura en Ciencias de la Computación, Universidad Nacional de Cuyo, Instituto Tecnológico Universitario (ITU), Universidad Nacional de Cuyo y la Universidad de Mendoza. Pueden consultarse más detalles de dichos laboratorios remotos y experiencias de uso por parte de estudiantes de las carreras mencionadas en [1,2,3,4,5,6,7].

1. INTRODUCCIÓN

1.1 Objetivos

En este proyecto se propone, como objetivos principales, adicionar algunos experimentos virtuales (o simulados) a las plataformas ya disponibles.

Como segundo objetivo, se espera colaborar con el avance de una tesis doctoral que tiene como título: "Desarrollo de competencias científico-tecnológicas en alumnos universitarios de carreras del área de las ciencias exactas mediante el método STEAM y el pensamiento computacional, necesarias para afrontar los desafíos del siglo XXI".

Como tercer objetivo, se espera obtener estadísticas de uso, opiniones, críticas y sugerencias por parte de estudiantes con la finalidad de mejorar estas herramientas como también en la forma de utilizarlas en clases y como herramientas de estudio por parte de los estudiantes.

1.2 Trabajos Relacionados

Existen un número importante de simulaciones y laboratorios virtuales accesibles a través de Internet. Se mencionan a continuación 3 ejemplos. Existen varios laboratorios virtuales de física, uno de los más importantes son las Simulaciones Interactivas de la Universidad de Colorado [8], en la cual se proveen laboratorios virtuales de varios tipos, por ejemplo, laboratorios de ondas, laboratorios de termodinámica y laboratorios de óptica física. La Universidad Técnica de Berlín provee un simulador de ventanas deslizantes de TCP ampliamente utilizado en cursos de redes de computadoras [9], [10]. El Instituto Tecnológico de Massachusetts provee un simulador de lenguaje ensamblador de 8 bits [11]. El mismo permite introducir a los estudiantes en el empleo de dicho lenguaje y de la arquitectura de procesadores.

1.3 Estado actual de avance del proyecto

El proyecto se encuentra en sus etapas finales de ejecución. Actualmente se han desarrolado tres laboratorios virtuales:

1 y 2) Laboratorio virtual de Efecto Doppler y laboratorio virtual de ondas mecánicas. Existen en la bibliografía algunos laboratorios virtuales de física muy completos, pero carecen de algunas experiencias de laboratorio necesarias por estudiantes que cursan Física 2. Se menciona que uno de los integrantes del grupo de trabajo del proyecto es profesor de Física 2 en la Facultad de Ciencias Exactas y Naturales de la Universidad Nacional de Cuyo.

3) Laboratorio virtual de ventanas deslizantes de redes de computadoras. Este es un tema visto en la materia Redes de Computadoras de la carrera Licenciatura en Ciencias de la Computación, como también en otras carreras de redes o computación. Existen en la bibliografía algunos simuladores, pero carecen de detalles profundos del comportamiento del mecanismo, siendo estos detalles importantes para los estudiantes de la carrera mencionada. Se menciona que uno de los integrantes del grupo de trabajo del proyecto es profesor de la asignatura Redes de Computadoras, en la Facultad de Ingeniería de la Universidad Nacional de Cuyo.

Los laboratorios virtuales a implementar están basados en tecnología web, de manera de poder ser accedidos a través de navegadores web típicos, desde cualquier computadora o teléfono celular. Las figuras 1, 2 y 3 muestran las interfaces web de estos laboratorios virtuales.

Figura 1: Interfaz de usuario del laboratorio virtual de efecto Doppler

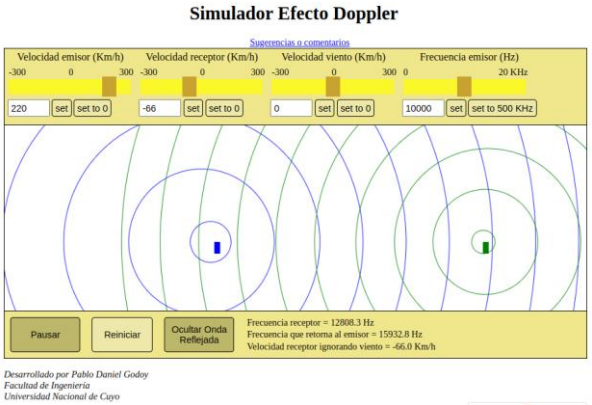


Figura 2: Interfaz de usuario del laboratorio virtual de ventanas deslizantes

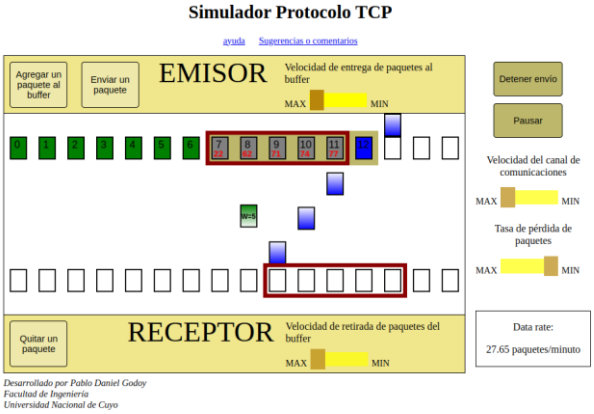
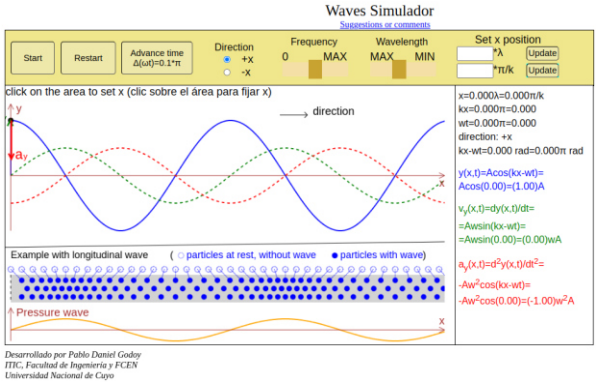


Figura 3: Interfaz de usuario del laboratorio virtual de ondas mecánicas



2. LINEAS DE INVESTIGACION y DESARROLLO

2.1 Laboratorios virtuales

Estos se basan en un modelo físico o de funcionamiento de diferentes tipos de sistemas implementado mediante software [11], [12]. Tienen como ventaja que pueden escalar más fácilmente y no hay limitación sobre la cantidad de equipamiento disponible, sobre todo si se implementan en el cloud [13]. Además, si están disponibles a través de Internet, pueden estar disponibles todos los días del año, las 24 horas. Tienen como desventajas que los modelos pueden no ser exactos y los estudiantes no trabajan sobre equipos ni sistemas reales. Sin embargo, su utilización ha demostrado a los investigadores que forman parte de este proyecto que pueden ayudar en gran medida a que sus estudiantes comprendan fácilmente conceptos de difícil entendimiento.

2.2 Prácticas educativas STEAM

Actualmente existe un amplio abanico de herramientas digitales que pueden usarse en la enseñanza de la ciencia, la tecnología, la ingeniería y la matemática (educación STEAM). Como parte del proyecto, una tesis doctoral en curso tiene como objetivo probar que una adecuada relación entre las prácticas educativas STEAM y las herramientas digitales puede servir para mejorar tanto las competencias científicas tecnológicas de los estudiantes como sus competencias digitales, necesarias para el desarrollo personal y profesional en la era digital. Para comprobar esta hipótesis, se han diseñado instrumentos con pautas para evaluar el desarrollo de competencias científicas, tecnológicas, digitales y las necesarias para afrontar los desafíos del siglo XXI, en alumnos pertenecientes a carreras del área de las ciencias exactas e ingeniería [14]. Por otro lado, este trabajo de tesis doctoral requiere de las herramientas para llevar adelante prácticas STEAM, entre estas, los simuladores.

3. RESULTADOS OBTENIDOS/ESPERADOS

3.1 Contar con laboratorios virtuales disponibles a través de Internet

Se cuenta actualmente con varios laboratorios virtuales destinados a educación. Se espera terminar de poner a punto estas herramientas, hacerlas disponibles de forma abierta a través de Internet y obtener estadísticas de uso y realimentación por parte de los estudiantes.

3.2 Obtener estadísticas de uso y retroalimentación de los estudiantes.

En esta etapa del proyecto se espera poner a prueba las herramientas implementadas durante el dictado de clases con estudiantes. Durante 2024 los laboratorios virtuales implementados fueron empleados para demostraciones en clases con los estudiantes, pero no se permitió el libre acceso de los estudiantes a los mismos, ni se recopilaron estadísticas de uso u opiniones de los estudiantes.

En esta etapa del proyecto los laboratorios virtuales serán utilizados durante el dictado de las materias Física 2, parte A y parte B, en la Facultad de Ciencias Exactas y Naturales de la Universidad Nacional de Cuyo y Redes de Computadoras, en la Facultad de Ingeniería de la Universidad Nacional de Cuyo. Se espera llevar a cabo las siguientes actividades:

a - Utilización para demostraciones en clases.

b - Permitir el acceso a los estudiantes para que puedan realizar actividades prácticas sobre los mismos.

c - Recopilar estadísticas de uso y obtener comentarios de retroaliemtación.

No se realizarán actividades que fuercen a los estudiantes a utilizar las herramientas, pues se desea analizar el uso de forma espontánea por parte de los estudiantes. Algunas de las métricas a analizar son:

a - Cantidad de estudiantes que acceden a los laboratorios virtuales luego de haber sido utilizados en clases para realizar demostraciones.

b - Tiempo de uso y acciones realizadas.

c - Valoración por parte de los estudiantes mediante encuestas anónimas.

d - Obtener comentarios, críticas y sugerencias.

3.3 Realizar aportes en las áreas de los laboratorios virtuales y enseñanza

El objetivo principal de este proyecto es desarrollar laboratorios virtuales y aplicarlos en el ámbito educativo. Se prevé que este proyecto genere conocimientos y resultados que sean de gran utilidad para la comunidad científica. Los hallazgos, los conocimientos adquiridos, las experiencias obtenidas en su aplicación docente y los resultados alcanzados serán documentados y presentados en congresos y publicados en revistas científicas. Además, los laboratorios virtuales desarrollados estarán disponibles de forma gratuita y abierta. Esto permitirá que profesores y estudiantes de cualquier institución educativa puedan utilizarlos.

4. FORMACIÓN DE RECURSOS HUMANOS

El director del proyecto realizó su doctorado en temas relacionados con laboratorios remotos, siento el título de su tesis: “Plataforma de desarrollo de laboratorios remotos de redes de sensores inalámbricos basados en cloud computing”. El presente trabajo es una evolución del trabajo que comenzó con dicha tesis doctoral.

Una de las integrantes del equipo de trabajo se encuentra realizando su doctorado. El título de la tesis doctoral es “Un modelo de enseñanza basado en el enfoque STEAM y el pensamiento computacional para fortalecer las competencias científicas de los alumnos de carreras del área de ciencias exactas”, estando dicha tesis fuertemente relacionada con el proyecto. Se espera que dicha tesis sea defendida durante el año 2025 o comienzos de 2026.

AGRADECIMIENTOS

Los autores agradecen el financiamiento recibido de la Secretaría de Investigación, Internacionales y Posgrado (SIIP),

Universidad Nacional de Cuyo, a través de los proyectos "SIIP B008: Desarrollo y aplicación de laboratorios virtuales destinados a educación de grado y pregrado" y “SIIP B036-T1: La seguridad de los dispositivos utilizados en IoT mediante la utilización de Blockchain, NFT y/o VPN”.

5. BIBLIOGRAFIA

[1] Godoy, Pablo; Marianetti, Osvaldo; García Garino, Carlos; "Experiences With Computer Architecture Remote Laboratories"; en Handbook of Research on Software Quality Innovation in Interactive Systems”, Chapter 5; IGI Global, Hershey, Pennsylvania, USA, 2020.

[2] Godoy, Pablo; Cayssials, Ricardo; García Garino, Carlos; "A WSN testbed for teaching purposes"; IEEE Latin America Transactions. 14(7),3351–3357 (2016).

[3] Godoy, Pablo; Cayssials, Ricardo; García Garino, Carlos; "Communication channel occupation and congestion in wireless sensor networks"; Computers & Electrical Engineering. 72, 846–858 (2018).

[4] Godoy, Pablo; Cayssials, Ricardo; García Garino, Carlos; "Laboratorio remoto para la formación de usuarios basado en el cloud"; Revista Iberoamericana de Educación en Tecnología y Tecnología en Educación 18, 7–18 (2016).

[5] Godoy, Pablo; García Garino, Carlos; Cayssials, Ricardo; "A nomadic testbed for teaching computer architecture"; In: XVII Workshop Tecnología Informática Aplicada en Educación (WTIAE), XXIV, CACIC, 2018.

[6] Godoy, Pablo; Cayssials, Ricardo; García Garino, Carlos; "Zigbee WSN Round Trip Latency in Function of Channel Occupation and Nodes Configuration"; In IEEE ArgenCon 2016.

[7] Godoy, Pablo Daniel; Marianetti, Osvaldo; Mansilla Roberto Alejandro; García Garino, Carlos; "Estudio de herramientas para la implementación de laboratorios remotos"; en XXIII Workshop de Investigadores en Ciencias de la Computación (WICC) 2021, Chilecito, La Rioja.

[8] Universidad de Colorado, PhET Interactive Simulations website. Accedido en febrero de 2022.

[9] Telecommunication Networks Group, Universidad Técnica de Berlín, simulador de ventanas deslizantes de TCP:

[10] López-Pedroza I.A., de Asís López-Fuentes F. (2021) Teaching of TCP Fundamental Operations Using a Digital Tool. In: Computer Science and Health Engineering in Health Services. COMPSE 2020. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 359. Springer, Cham.

[11] Darrah, M., Humbert, R., Finstein, J. et al. Are Virtual Labs as Effective as Hands-on Labs for Undergraduate Physics? A Comparative Study at Two Major Universities. J Sci Educ Technol 23, 803–814 (2014).

[12] Dongfeng Liu, Priscila Valdiviezo-Díaz, Guido Riofrio, Yi-Meng Sun, Rodrigo Barba, Integration of Virtual Labs into Science E-learning, Procedia Computer Science, Volume 75, 2015, Pages 95-102, ISSN 1877-0509.

[13] José A. González-Martínez, Miguel L. Bote-Lorenzo, Eduardo Gómez-Sánchez, Rafael Cano-Parra, Cloud computing and education: A state-of-the-art survey, Computers & Education, Volume 80, 2015.

[14] Manganelli, Silvina. Aportes de las herramientas digitales a STEM, durante la investigación científica, para fomentar el desarrollo de competencias científicas, tecnológicas y digitales. Congreso Argentino de Ciencias de la Computación (CACIC).

Tecnología Informática Aplicada en Educación (WTIAE). 2021.

Aprendizaje Personalizado y Colaborativo
El Potencial de la Tecno pedagogía en la Formación Universitaria

Mg. Roberto Bertone[1], Mg. José Luis Filippi[2], Lic. Guillermo Lafuente[3], Mg. Carlos Ballesteros[4], Lic. Gustavo Lafuente[5], Pablo Etcheverry[6], Valentín Filippi[7]
LIAU[9] - Facultad de Ingeniería – UNLPam.
pbertone@ada.info.unlp.edu.ar[1]
{filippij[2], lafuente[3], balleste[4], gustavo[5], etcheverry[6]}@ing.unlpam.edu.ar
valefilippi08@gmail.com[7]
[9]Laboratorio de Investigación de Ambientes Ubicuos

Resumen

La tecno pedagogía es la tecnología aplicada a la educación. A través de ella se busca dar respuesta pedagógica al impacto de las herramientas digitales en el proceso formativo.

La tecno pedagogía utiliza estrategias didácticas como el aprendizaje significativo y colaborativo, conjuntamente incorpora recursos tecnológicos permitiendo el desarrollo de actividades de enseñanza virtuales que permitirán a los estudiantes accionar de forma autónoma.

Los centros educativos del siglo XXI se han de enfrentar a las nuevas demandas de una sociedad cambiante y con una clara vocación tecnológica. Aparecen nuevas formas de abordar lo tecnológico: aprendizaje adaptativo y móvil, inteligencia artificial, realidad virtual y aumentada y surgen las interfaces de uso natural (Johnson et al., 2016; Becker et al., 2017). [1]

El propósito del proyecto, de naturaleza teórico-práctico, consiste en indagar las posibilidades que ofrecen los dispositivos tecnológicos de última generación, como instrumentos aplicados a diferentes situaciones de aprendizaje, considerando los aspectos pedagógicos y las teorías del aprendizaje como marcos flexibles de la organización pedagógica. Las actividades que se proponen tienen lugar en el contexto de la UNLPam.

Palabras claves: Tecno pedagogía. Aprendizaje Móvil. Aprendizaje Colaborativo.

Contexto

Tipo de Investigación: Aplicada

Campo de Aplicación Principal: 7 1802 Computación, 7 1803 Comunicaciones.

Campos de Aplicación posibles: 13 1040 Ciencia y Tecnología, 7 4399 Otras – Educación – Tecnología Aplicada a la Educación

Institución que Coordina el Proyecto: Facultad de Ingeniería, Universidad Nacional de La Pampa.

Introducción

Con la aparición del coronavirus a comienzos del 2020, los cambios tecno pedagógicos que se realizaban de manera natural en los centros educativos, se precipitaron como resultado de la situación de encierro ciudadano.

A partir del 2022 con la vuelta a la formación presencial, se visualiza un modelo educativo donde lo presencial y lo virtual van a convivir de forma permanente, y en este punto el accionar docente es vital para desplegar su actividad profesional según los modelos que la sociedad moderna exige.

Aparecen nuevas estrategias formativas como el microlearning **[2]** y el aprendizaje inversivo. Surge el desafío educativo, definir metodologías y procesos para desplegar una formación pedagógica eficaz. Por ello se requiere un diseño tecno pedagógico instruccional adaptado al contexto en el que se va a desplegar la tarea educativa, con

docentes capacitados en desempeñar su profesionalidad de forma hibrida. **[3].**

A partir del estrepitoso cambio de paradigma educativo que se articuló en marzo del 2020, donde se determina el paso de la formación presencial a la virtual, numerosos actores se pronuncian en priorizar el desarrollo de habilidades integrales además de las cognitivas. Por ello es necesario "…construir otras formas de hacer educación, y que los cambios curriculares den cuenta de la posibilidad, la necesidad y la urgencia de vincular los contenidos disciplinares con habilidades de aprendizaje socioemocionales en las aulas…". **[4]**

Aquí radica el objetivo primario del proyecto: indagar las posibilidades que ofrecen los dispositivos tecnológicos de última generación, como instrumentos aplicados a diferentes situaciones de aprendizaje, considerando los aspectos pedagógicos y las teorías del aprendizaje como marcos flexibles de la organización pedagógica.

Situación Actual del Problema.

La educación influenciada por ésta realidad, acepta la necesidad de reconfigurar el proceso educativo, partiendo de algunas premisas:

a. Redefinir nuevas acciones para docentes y estudiantes.

El creciente número de dispositivos móviles y la conectividad a internet de forma permanente, permite a los estudiantes adquirir nuevos contenidos en línea y configurar sus conocimientos a partir de intereses propios. El estudiante es artífice de su futuro, asume un papel activo, decide que aprender y cómo.

El rol tradicional del docente es superado por nuevas exigencias, requiere nuevas capacidades y conocimientos para desempeñar su actividad en entornos virtuales de aprendizaje.

Frente a la realidad del mundo digital que rodea a los estudiantes de hoy en día, es justo que los docentes replanteen su rol en la construcción del conocimiento sobre ellos. De allí que la competencia exigida a un profesor del Siglo XXI es precisamente que se enfoquen "en las necesidades de los alumnos, supervisando su búsqueda de información e intentando facilitar la búsqueda de información individual de los alumnos ya que el papel de suministrador de conocimiento ha sido superado por las TIC" **[5]**.

Mucho se habla de cuáles serían las competencias digitales para un docente, un acercamiento a ello lo propone Núñez-Torrón Stock **[6]**, quien plantea cinco competencias básicas para el docente en la educación, y entre ellas se pueden destacar dos:

1. Usar herramientas de trabajo en línea.
2. Utilizar dispositivos móviles en el aula.

En la mayoría de las universidades de todo el mundo se observa la incorporación de los sistemas de Gestión de Aprendizaje (LMS), para implementar entornos virtuales que permiten la gestión del aprendizaje. **[7]** Sin embargo algunas investigaciones indican que las motivaciones de los estudiantes son vistas como una obligación al participar en entornos de aprendizaje institucionales cerrados, mientras que la motivación es mayor cuando el proceso formativo es abierto a diferentes tecnologías en contextos informales. Estos espacios son conocidos como PLE (Personal Learning Environments) donde los estudiantes configuran su propio entorno. **[8]**

Algunos autores ya incorporan un nuevo concepto, el de mPLE (Mobile Personal Learning Environments) como una nueva estructura de aprendizaje para la generación que hace uso de los dispositivos móviles. **[9]**

La agregación del modelo tecno pedagógico, que integra las tecnologías en todas las áreas del conocimiento, va a permitir desplegar las competencias digitales de los docentes en el aula, influenciando de manera positiva el desempeño de los estudiantes.

La tecno pedagogía utilizada como herramienta de planificación pedagógica expone diferentes actividades formativas que

conectan lo que se enseña en la práctica con lo que aprende el estudiante.

Es un modelo que viabiliza la construcción del conocimiento de forma colaborativa, donde los docentes efectúan su aporte según su área del conocimiento específico, tecnología, pedagogía y/o contenidos. Así, los docentes aprenden unos de otros, lo que puede derivar en el desarrollo colaborativo del currículo. **[10]**

La organización de la tecno pedagogía en el aula, integra las tecnologías con la pedagogía y el conocimiento científico. Para que la integración sea efectiva, los docentes no solo deben poseer conocimientos en tecnología y pedagogía, se debe poseer además el conocimiento propio de la disciplina, y desde allí integrar el conocimiento tecnológico y el conocimiento pedagógico. **[11]**.

b. Facilitar la continuidad del aprendizaje.

El aprendizaje móvil permitirá a los estudiantes dar continuidad a su formación a lo largo de toda su vida.

Los estudiantes universitarios no solo deben dominar las disciplinas correspondientes a su carrera, además deben incorporar habilidades transversales como el pensamiento crítico, resolución de problemas, persistencia y trabajo colaborativo. **[12]**

En la actualidad se está transitando tiempos de pos pandemia, y si algo bueno se puede rescatar de esta situación, es la consolidación del uso de los dispositivos tecnológicos de última generación durante los últimos dos años, escenario que alcanzo a todos los sectores sociales incluidas las instituciones educativas en sus diferentes niveles. Entre los instrumentos tecnológicos que sobresalieron en las actividades formativas virtuales que se desplegaron durante la pandemia y que van a prevalecer pos pandemia, podemos indicar **[13]**, aulas colaborativas, educación personalizada e integrar en las actividades pedagógicas las tecnologías emergentes como nuevos ambientes de enseñanza y aprendizaje. Se trata no solo de afianzar los conocimientos adquiridos de los estudiantes, sino el de desarrollar en ellos las habilidades cognitivas y meta cognitivas para la realización de tareas en su quehacer diario, objetivo primario del diseño tecno pedagógico. **[14]**

c. Grandes volúmenes de datos. Big data.

Los investigadores que estudian el aprendizaje en línea, los sistemas de tutoría inteligente, los laboratorios virtuales, las simulaciones y los sistemas de gestión del aprendizaje están explorando maneras de entender y utilizar mejor la analítica del aprendizaje, a fin de mejorar la actividad del docente y seguir avanzando en la educación para todos. **[15]**

En éste contexto el punto de partida debe ser siempre tecno pedagógico. No se trata de innovar por innovar, hay que explotar el potencial que ofrecen las nuevas herramientas tecnológicas que hacen posible el aprendizaje móvil, a partir de una planificación docente previa dentro de un marco teórico formalizado.

El BigData trae beneficios de cara a la personalización de la educación en el alumno/a de modo que estudiantes como profesores sepan en qué punto del aprendizaje se encuentran, cuáles son sus debilidades y fortalezas. **[16]**

Línea de Investigación y Desarrollo

El plan de actividades corresponde al proceso de investigación aplicada, con objetivos de corto, mediano y largo plazo, y una duración prevista de cuatro años.

<u>Primer año.</u>

- Identificar el porcentaje de inserción de dispositivos móviles en la institución (teléfonos y/o tablets).
- Analizar la inserción del aprendizaje móvil en las universidades de todo el mundo.

- Estudiar como el aprendizaje móvil puede cerrar la brecha entre el aprendizaje formal y el informal.
- Identificar los escenarios educativos que muestren aspectos móviles en las actividades formativas del ámbito académico propio.
- Examinar herramientas tecnológicas móviles orientadas al proceso educativo de acceso libre y gratuito.

Segundo y Tercer año.

- Definir el enfoque tecno pedagógico que posibilite un aprendizaje móvil de calidad.
- Estudiar las tecnologías disponibles para implementar acciones de aprendizaje móvil en el contexto educativo universitario.
- Definir prioridades según necesidades de inmediatez en la virtualización de las disciplinas que se han de impartir.
- Gestionar el uso de plataformas de formación virtual que posibilite una doble modalidad, presencial y mediada por las nuevas tecnologías móviles.
- Incorporar el uso de redes sociales que favorezcan la práctica educativa.
- Confeccionar objetos de aprendizaje acorde a las herramientas disponibles en la nube y a las características de los dispositivos móviles que van a operar el producto final.
- Diseñar y/o desarrollar aplicaciones móviles a partir de las necesidades que se presenten durante el transcurso del proceso educativo virtual.
- Capacitar a la comunidad educativa en general en el desarrollo de objetos de aprendizaje a partir de los requerimientos de carácter institucional.
- Difundir los avances a toda la comunidad universitaria los progresos a medida que van trascurriendo, a través de jornadas, congresos y/o revistas científicas.
- Propiciar el intercambio de información permanente con grupos de investigación que den valor agregado a nuestra actividad profesional.

Cuarto año.

- Instituir las aplicaciones desarrolladas para dispositivos móviles que hayan alcanzado buen nivel de aceptación.
- Conformar un repositorio de objetos de aprendizajes de acceso libre.
- Registrar y difundir los resultados alcanzados con la finalidad de que se puedan utilizar, ampliar y mejorar a través de trabajos futuros.
- Disponer de un catálogo de trabajos realizados en el sitio web del grupo de investigación GIAU (Grupo de Investigación de Ambientes Ubicuos) a partir del cual los interesados podrán acceder al material requerido.
- Presentar en jornadas, congresos y/o revistas de todo el mundo los resultados alcanzados.

Resultados Obtenidos/Esperados

El proyecto da inicio a su actividad de investigación y desarrollo durante el 2023.

Durante el 2024 se llevaron a cabo varios trabajos destacados, que se presentaron en diferentes reuniones científicas; a modo de ejemplo se pueden citar: **SENIE** 2024, se presentó la incorporación de Micro Aprendizajes en el aula. **CONAIISI** 2024. se expuso sobre un caso de estudio sobre la implementación de una aplicación de Simulador UNOARDUSIM. **IDETEC** 2024, se exhibió la aplicación UNLPam móvil que conecta la comunidad universitaria transformando la experiencia educativa. **CIIDDI** 2024, se expuso sobre la Tecnología H5P y aprendizaje auto dirigido. **WITE** 2024, se presentó un análisis sobre las competencias digitales en estudiantes de ingeniería.

Formación de Recursos Humanos

Director de Proyecto

Co-Director de Proyecto

6 Investigadores

Referencias

[1] Integración de Tecnologías de la Información y la Comunicación a la Docencia - 2017 https://nuestroscursos.net/mod/url/view.php?id=1857

[2] Que es un Nanodegree. Webinar. 9 de febrero de 2016 https://www.itmadrid.com/que-es-un-nanodegree/
[3] Tendencias eLearning 2022. https://www.itmadrid.com/wp-content/uploads/2021/12/itmadrid-tendencias-elearning-2022.pdf
[4] [UNESCO (2021), Educación pos pandemia: salud mental y prevención de la violencia con la comunidad educativa, Memoria, UNESCO, en: UNESCO (2021), Educación pos pandemia: salud mental y prevención de la violencia con la comunidad educativa, Memoria, UNESCO, en: https://es.unesco.org/sites/default/files/doc-memoria-foro-educacion-post-pandemia.pdf.
[5] webdelmaestrocmf.com, "Estas son las capacidades digitales de un docente moderno." [Online]. Available: http://webdelmaestrocmf.com/portal/5-competencias-digitales-que-deben-tener-los-profesores-actuales/?utm_source=blogsterapp&utm_medium=Facebook. [Accessed: 27-Oct-2017].
[6] A. Núñez-Torrón Stock, "5 competencias digitales que deben tener los profesores actuales." [Online]. Available: http://www.ticbeat.com/educacion/5-competencias-digitales-profesores/. [Accessed: 27-Oct-2017].
[7] M. P. Prendes, "Plataformas de campus virtual con herramientas de software libre: Análisis comparativo de la situación actual en las universidades españolas," Murcia, Project Report EA-2008-0257, Jun. 2009. [Online]. Available: http://www.um.es/campusvirtuales/informe_final_CVSL_SF.pdf
[8] F. J. García-Peñalvo, "Docencia," in Libro blanco univ. digital 2010, J. Laviña Orueta and L. Mengual Pavón, Eds. Barcelona, Spain: Ariel, 2010, pp. 29-61.
[9] G. Attwell, J. Cook, and A. Ravenscroft, "Appropriating Technologies for Contextual Knowledge: Mobile Personal Learning Environments," in Best Practices for the Knowledge Society. Knowledge, Learning, Development and Technology for All, M. D. Lytras, P. O. de Pablos, E. Damiani, D. Avison, A. Naeve, and D. G. Horner, Eds. Springer Berlin Heidelberg, 2009, pp. 15-25.
[10] Technological Pedagogical Content Knowledge (TPACK) in Action. https://www.researchgate.net/publication/271937971_Technological_Pedagogical_Content_Knowledge_TPACK_in_Action
[11] [Kiray, S. A., Çelik, I., & Çolakoğlu, M. H. (2018).TPACK Self-efficacy perceptions of science teachers:A structural equation modeling study. Egitim ve Bi-lim, 43(195), 253-268. http://egitimvebilim.ted.org.tr/index.php/EB/article/view/7538]
[12] World Economic Forum, 2015.
[13] Tendencias educativas que marcaran el 2022 y el futuro de las aulas. https://www.entreestudiantes.com/2022/01/tendencias-educativas-marcaran-2022-futuro-aulas/.
[14] Pedroza, O., & Crespo, M. (2017). Importancia del diseño tecno pedagógico basado en el enfoque de la acción, para reforzar el dominio del idioma inglés como segunda lengua. Revista Colombiana de Computación, 18(2), 7-21, e-ISSN: 2539-2115. Recuperado de: https://doi.org/10.29375/25392115.3214
[15] Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing teaching and learning through educational data mining and learning analytics: An issue brief.US Department of Education, Office of Educational Technology, 1-57.
[16] Otero, A. (2020). El Big Data. Como se usaría en la educación, vermislab. Recuperado de https://www.vermislab.com/big-data-que-es-como-nos-afecta-y-su-uso-en-educacion/

Avances en la elaboración de estrategias para la inclusión de la Ciencia de Datos en la escuela secundaria

Sofía Martin[1], Claudia Banchoff[1], Paula Venosa[1], Liliana Hurtado[2]
[1]LINTI - Laboratorio de Investigación en Nuevas Tecnologías Informáticas.
Facultad de Informática. Universidad Nacional de La Plata
Calle 50 esq. 120, 2do Piso. Tel: +54 221 4223528
{smartin, cbanchoff, pvenosa}@info.unlp.edu.ar}
[2]Facultad de Ciencias Económicas, Jurídicas y Sociales. Universidad Nacional de Salta
lilianahur@hotmail.com

RESUMEN

En este trabajo se describe una línea de investigación trabajada desde el año 2022 focalizada en analizar, llevar adelante y evaluar propuestas pedagógicas que permiten introducir aspectos básicos de la Ciencia de Datos en escuelas secundarias. El enfoque propuesto es aprovechar las distintas iniciativas que ya se implementan en las escuelas, donde se enseñan contenidos de la Ciencia de la Computación, vinculándolo con estrategias para realizar análisis de datos.
Se hace hincapié en el uso de herramientas de software libre y datos abiertos.
Se plantean el contexto de esta investigación, el estado de avance y los resultados obtenidos hasta el momento.

Palabras clave: Ciencia de Datos, educación secundaria, Python, datos abiertos, programación de computadoras

CONTEXTO

En los últimos tres años se han implementado una serie de experiencias y actividades que permiten definir y evaluar distintas estrategias para introducir aspectos básicos de la Ciencia de Datos en el ámbito de la escuela secundaria. Esta ciencia define como un campo de estudio que permite planificar, adquirir y extraer conocimientos significativos a partir de los datos [1]. Se compone de la intersección de: la Ciencia de la Computación, las Estadísticas y la disciplina específica que describe el dominio del problema, siendo su perspectiva interdisciplinar y relacional clave para realizar un análisis significativo [2]. Estas experiencias se sustentan en la larga trayectoria que cuenta este equipo de investigación en el campo, especialmente enfocados en trabajar distintas formas de incorporar el pensamiento computacional a través de proyectos de robótica [3] y otras propuestas similares [4].
Aunque las temáticas varían, los desafíos de la implementación persisten: adaptar los contenidos para integrarlos a la currícula y adecuar los recursos informáticos al nivel educativo para garantizar una efectiva implementación en el aula [5]. Estas problemáticas continúan siendo áreas de interés y vacancia en la enseñanza de la Ciencia de Datos, una disciplina que aún no ha sido ampliamente adoptada en las escuelas, especialmente en nuestra región.
La relevancia de la Ciencia de Datos y la búsqueda de estrategias pedagógicas e informáticas para su adecuada adaptación al nivel educativo correspondiente, continúan siendo el foco de atención de esta línea de investigación, llevada adelante por docentes-investigadores del LINTI. Esta línea de trabajo se enmarca en el proyecto: “Tecnologías digitales para la Inclusión, la Equidad y la Sostenibilidad", del Programa Nacional de Incentivos, y, a través de la ejecución de dos proyectos de extensión presentados en convocatorias del programa Universidad, Cultura y Territorio de la Secretaría de Políticas Universitarias (SPU), del Ministerio de Educación de la Nación Argentina, se ha podido evaluar la aplicación de las citadas propuestas.

1. INTRODUCCIÓN

La Ciencia de Datos, que ha cobrado relevancia en los últimos años, debido a su amplia gama de posibles aplicaciones [6], aún no ha sido introducida como tal en la escuela secundaria. Si bien se abordan aspectos básicos de análisis de datos en asignaturas como matemáticas, por ejemplo, este abordaje está

orientado a la aplicación de cálculos estadísticos sin una conexión clara con otras áreas del conocimiento.

La introducción de aspectos básicos de la Ciencia de Datos en la currícula de nivel secundario da lugar a que los estudiantes puedan desarrollar habilidades y competencias para conectar sus aprendizajes con el mundo real, fomentando el pensamiento crítico y la capacidad para la resolución de problemas en un entorno digital impulsado por los datos [7]. Permite formular desafíos relacionados con problemáticas del mundo actual utilizando **datos abiertos**, lo cual también promueve la transparencia en un contexto donde la información a menudo es comunicada con sesgos y donde interpretarla correctamente se vuelve crucial. Esto permite abordar así algunos aspectos éticos en el ámbito educativo [8].

Numerosos repositorios gubernamentales y públicos ofrecen datasets de contenidos variados. En nuestros talleres, se focaliza en datasets que contienen información de recursos naturales de nuestro país, como ser lagos, áreas protegidas y fauna argentina. Esto también permite analizar y observar algunas características destacadas del país y difundir los repositorios que contienen estos datos. También se incorporaron conjuntos de datos sobre música, películas o deportes, buscando integrar información relevante y local en estos temas.

La ludificación de las actividades en los talleres de Ciencia de Datos para estudiantes de secundaria se presenta como una estrategia efectiva para fomentar la motivación y la participación activa. Se plantean preguntas desafiantes que requieren que los estudiantes apliquen los conceptos aprendidos sobre análisis y filtrado de datos en los conjuntos de datos proporcionados. Este enfoque no solo estimula el pensamiento crítico, sino que también permite a los estudiantes experimentar de manera práctica los principios de la Ciencia de Datos, haciendo que el aprendizaje sea dinámico y motivador.

Además, se promueve el trabajo en grupo como una herramienta para potenciar las habilidades individuales de los estudiantes. Al colaborar en la resolución de problemas y en la interpretación de los datos, los alumnos no solo desarrollan competencias técnicas, sino que también mejoran sus habilidades interpersonales y de comunicación.

Como se mencionó anteriormente, las líneas de trabajo se centran en abordar los dos desafíos principales: la adecuación de los contenidos al nivel educativo y la elección de las herramientas más adecuadas para la realización de las actividades.

A partir de experiencias documentadas sobre la enseñanza de la Ciencia de Datos [9][10], se identifican aspectos clave que también se reflejan en nuestras experiencias. Entre ellos, destacan la dificultad en la búsqueda de datos específicos en los repositorios públicos por parte de docentes no especializados, la resistencia de algunos educadores a incorporar contenidos nuevos, para los cuales no recibieron formación durante su preparación profesional, y la necesidad de adaptar tanto los contenidos como las herramientas informáticas al contexto educativo. Además, en varias de estas experiencias, se resalta la importancia de trabajar con datos abiertos como un recurso valioso para facilitar el aprendizaje y concretar análisis de datos reales.

Las primeras experiencias se centraron en trabajar con situaciones semi resueltas utilizando, Jupyter Notebook[1], una herramienta ampliamente utilizada [11], donde las soluciones implican la programación en Python utilizando la librerías pandas[2] [12]. Para la comunicación de los resultados se utilizan las librerías Matplotlib[3] y Plotly[4] y en algunos casos más avanzados la herramienta Streamlit[5], que permite la generación de una aplicación web y su posterior alojamiento en un repositorio público de fácil acceso.

Algunas dificultades surgidas en aquellas instituciones cuyos docentes no cuentan con

[1] https://jupyter.org/

[2] https://pandas.pydata.org/

[3] https://matplotlib.org/

[4] https://plotly.com/python/

[5] https://streamlit.io/

formación en programación, dio lugar al estudio y evaluación de soluciones alternativas que simplifiquen el aprendizaje de las instrucciones programáticas como es el caso de la programación en bloques [13].

2. LÍNEAS DE INVESTIGACIÓN, DESARROLLO E INNOVACIÓN

Esta línea de investigación se centra en los siguientes temas:

- Análisis de las habilidades adquiridas por los estudiantes de nivel secundario al introducir aspectos básicos de la Ciencia de Datos.
- Relevamiento de experiencias de enseñanza de la Ciencia de Datos en escuelas de distintas regiones, incluyendo metodologías y herramientas utilizadas, con foco en el uso de software libre y datos abiertos.
- Evaluación y desarrollo de herramientas informáticas que permitan el desarrollo de las distintas actividades propuestas.
- Diseño, elaboración y evaluación de materiales didácticos para la enseñanza de la Ciencia de Datos en el nivel secundario.
- Diseño, ejecución y evaluación de actividades específicas con estudiantes y docentes de nivel secundario.

3. RESULTADOS Y OBJETIVOS

Las experiencias realizadas han dejado algunos resultados interesantes.

Respecto a los conjuntos de datos y propuestas de actividades:

- El abordaje de la limpieza de datos mediante un proceso guiado, donde se trabaja con conjuntos de datos pre-analizados que incluyen algunos datos "erróneos conocidos", facilita la comprobación de que los datos sean completos y correctos.
- Durante la búsqueda de conjuntos de datos motivadores, se fomenta la generación de datos propios, lo que invita a abordar problemáticas específicas de la escuela o comunidad y, al mismo tiempo, permite aprender a diseñar y realizar encuestas relacionadas con dichas problemáticas.
- Se presentan distintas formas de comunicar los resultados de los análisis, lo que permite que los estudiantes analicen la efectividad de cada opción.
- Una adecuada elección de los datos permite reflexionar sobre aspectos éticos en la comunicación de los análisis, como la introducción de sesgos o el hallazgo de datos erróneos que no permiten la verificación de los resultados obtenidos, por ejemplo.

Respecto a las herramientas informáticas utilizadas:

- Se elaboraron una serie de criterios que permitieron evaluar herramientas alternativas a aquellas que requieren programación directa en Python, la cual fue utilizada inicialmente [13].
- Se descartaron algunas de las herramientas analizadas inicialmente, ya que el objetivo es la vinculación con la programación de computadoras, lo que implica la necesidad de contar con herramientas que permitan aplicar distintas estrategias didácticas para abordar la programación de las soluciones. Por ello, se analizaron especialmente aquellas que permiten la programación en bloques [14].
- Se decidió incluir la librería Folium[6], en en los entornos Jupyter Notebook o Streamlit, debido a que permite la visualización de datos georeferenciados utilizando los mapas como otro recurso de visualización.

De acuerdo a la evaluación realizada de herramientas y considerando que ninguna contaba con todas las características deseadas, se desarrolló en el marco de una tesina de grado, la herramienta **Pandalyze**[7] [15], una aplicación web de código abierto, diseñada especialmente para introducir aspectos básicos de análisis de datos en la escuela secundaria. En particular, esta herramienta ofrece una interfaz de programación basada en bloques que incluye alguna de la funcionalidad provista por la librería pandas de Python. Los bloques incluidos en Pandalyze surgen de las actividades realizadas con docentes y

[6] https://python-visualization.github.io/folium/latest/

[7] Disponible en https://pandas.linti.unlp.edu.ar/

estudiantes. Es decir, que se incluyó un conjunto mínimo de bloques correspondiente a las funciones básicas de pandas y Python de manera tal de introducirlas de forma simplificada.

Pandalyze fue utilizada en dos talleres con estudiantes y docentes con resultados alentadores. Estos talleres se realizaron en el marco de los proyectos de extensión "Ciencia de Datos en la escuela" durante el año 2024.

Si bien se buscaron datos que incluyeran geolocalización para así poder utilizar mapas, ésta es una característica que aún no posee Pandalize. Actualmente se encuentra en desarrollo una actualización de la misma para extender ésta y otras funcionalidades adicionales.

Por otro lado, se realizaron experiencias con la herramienta Streamlit, un framework desarrollado en Python que permite generar una aplicación web de forma sencilla, con foco en la visualización de datos. Esta herramienta, que se proponen en contextos donde los participantes tienen conocimientos más avanzados de programación, y ha demostrado ser motivadora, ya que permite elaborar una aplicación web con interfaces actuales.

Respecto al **objetivo principal de esta línea de investigación** el mismo continúa siendo el mismo: estudiar y formular propuestas didácticas para la incorporación de aspectos básicos de la Ciencia de Datos, conforme a los lineamientos de las escuelas secundarias de la Provincia de Buenos Aires, con énfasis en el uso de herramientas de licencia libre y de datos abiertos. Estas propuestas involucran también la elección de las herramientas informáticas más adecuadas según el contexto de aplicación.

Para alcanzar este objetivo, se proponen los siguientes **objetivos específicos**:

- Actualizar el relevamiento de las experiencias pedagógicas en la región para la enseñanza de contenidos relacionados con las Ciencias de la Computación en escuelas secundarias, debido a la estrecha vinculación con la Ciencia de Datos. Esto podría permitir sacar provecho de las mismas para la introducción de nuevas propuestas.
- Evaluar las metodologías y herramientas informáticas utilizadas en la enseñanza de contenidos específicos de la Ciencias de la Computación y Ciencia de Datos, en las escuelas de nivel secundario.
- Evaluar librerías para el trabajo con Python, similares o que complementen a la librería Streamlit.
- Realizar una evaluación más exhaustiva de la herramienta Pandalize. desarrollada en el marco de esta línea de trabajo, y extender su funcionalidad para la incorporación de opciones de visualización de datos geográficos.
- Elaborar nuevas propuestas pedagógicas para la inclusión de aspectos básicos de la Ciencia de Datos en la escuela secundaria que incluyan el uso de Pandalyze.
- Diseñar, implementar y evaluar las intervenciones con docentes y estudiantes de nivel secundario.
- Actualizar el sitio web público[8] de referencia con el que se cuenta, incorporando ejemplos de análisis de datos que utilicen la visualización de mapas interactivos.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo de la línea de I+D+i presentada en este artículo está formado por docentes investigadores categorizados del LINTI y estudiantes de la Facultad de Informática, con trayectoria en el trabajo con escuelas [3] [4].

En particular, esta línea de investigación tiene varios años de trabajo y fue presentada en las ediciones 2023 y 2024 de este Workshop. En este contexto se desarrolló la tesina de grado denominada "**Pandalyze: aprende Ciencia de Datos con programación en bloques",** la cual fue defendida en noviembre de 2024. Así mismo, también se encuentra en desarrollo la tesis de maestría titulada "**Espacio de aprendizaje de la Ciencia de Datos para establecimientos educativos de la Provincia**

[8] Sitio público con recursos elaborados: https://ciencia_datos_escuela.gitlab.io/

de Buenos Aires", en la que las autoras de este trabajo están profundizando la investigación y desarrollo de estrategias para la introducción de aspectos básicos de la Ciencia de Datos en el contexto educativo provincial de nivel secundario.

Cabe destacar que esta línea de trabajo también se aborda en la asignatura "Seminario de Lenguajes, opción Python", lo cual ha motivado que grupos de estudiantes de grado participen en proyectos de extensión e innovación, ampliando los contenidos vistos. Así mismo, también se han incluido estudiantes de nivel medio de la EET 9 de La Plata que se encuentran desarrollando sus prácticas profesionales supervisadas (PPS) en el LINTI. Estas experiencias han proporcionado evidencia empírica para continuar indagando sobre las formas de incorporar esta disciplina en la educación media.

5. BIBLIOGRAFÍA

[1] Donoho, D. (2017) 50 Years of Data Science. Journal of Computational and Graphical Statistics 26(4), 745–766.

[2]Danyluk, A., & Leidig, P. Computing Competencies for Undergraduate Data Science Curricula: Vol. ACM Data Science Task Force. Association for Computing Machinery.

[3] Banchoff Tzancoff, C. M., Martin, E. S., Gómez, N. S., & López, F. E. M. (2019) Experiencias en robótica educativa: Diez años trabajando con escuelas argentinas. XIV TE&ET. ISBN 978-987-733-196-7, pág. 84-93.

[4] Queiruga C., Banchoff Tzancoff C, Martin S. y Kimura I. (2024) Propuestas de acercamiento a la Informática para la escuela secundaria: contenidos, estrategias didácticas y materiales 53 JAIIO - SAEI - ISSN: 2451-7496. pp. 116-134.

[5] Luxton-Reilly, (2016) A. Learning to Program is Easy. ACM Conference on Innovation and Technology in Computer Science Education. pp. 284-289.

[6] Hicks, S. C., & Irizarry, R. A. (2018) A Guide to Teaching Data Science. The American Statistician, Vol 72. Article 4.

[7] Rao, A. R., Desai, Y., & Mishra, K. (2019). Data science education through education data: An end-to-end perspective. IEEE Integrated STEM Education Conference (ISEC), 300-307.

[8] De Veaux, R. D., Agarwal, M., Averett, M., Baumer, B. S., Bray, A., Bressoud, T. C., ... & Ye, P. (2017). Curriculum guidelines for undergraduate programs in data science. Annual Review of Statistics and Its Application, 4(1), 15-30.

[9] Heinemann, B., Opel, S., Budde, L., Schulte, C., Frischemeier, D., Biehler, R., Podworny, S., & Wassong, T. (2018) Drafting a Data Science Curriculum for Secondary Schools. 18th Koli Calling International Conference on Computing Education Research, 1-5.

[10] Atenas, J., & Havemann, L. (2015) Open Data as Open Educational Resources: Case Studies of Emerging Practice.

[11] Randles BM, Pasquetto IV, Golshan MS, Borgman CL. (2017) Using the Jupyter Notebook as a Tool for Open Science: An Empirical Study. ACM/IEEE Joint Conference on Digital Libraries (JCDL)

[12] Martin S., Banchoff Tzancoff C., Venosa P. y Hurtado L. (2014) Ciencia de Datos en la escuela: herramientas y estrategias aplicadas en experiencias en La Plata. XIX TE&ET. ISBN 978-987-3977-95-4. pp 56-65.

[13] Martin S., Banchoff Tzancoff C., Marcón P. y Duggan T. (2014) Herramientas Informáticas utilizadas para Introducir Aspectos Básicos de Ciencia de Datos en la Escuela Secundaria. XXX CACIC. ISBN 978-950-34-2428-5. pp 350-359.

[14] Martin S., Banchoff Tzancoff C., Venosa P. y Hurtado L. (2024) Estrategias para incorporar análisis de datos en la escuela secundaria. XXVI WICC. ISBN 978-987-8352-57-2. pp 846-851

[15] Marcón P., Duggan T. (2024) Pandalyze: aprender Ciencia de Datos con programación en bloques. http://sedici.unlp.edu.ar/handle/10915/173859

Creación de tecnologías digitales para la innovación educativa

Sanz Cecilia[1,2][0000-0002-9471-0008] , Artola Verónica[1][0000-0003-0833-1077], Salazar Mesía Natalí[1][0000-0003-3946-8383], Iglesias Luciano[1][10000-0002-3284-6997], Buffarini Abril[1], Ibañez Bárbara[1], Bruno Laureano[1], Zeballos Matías[1], Del Gener Aldana[1][0009-0008-0024-3703], Segovia Maite[1], Astudillo Gustavo[3], Baldassarri Sandra[4][0000-0002-9315-6391]

[1]Instituto de Investigación en Informática LIDI (III-LIDI). Centro Asociado CIC.
Facultad de Informática – Universidad Nacional de La Plata
[2] Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC)
[3] Facultad de Ciencias Exactas y Naturales - Universidad Nacional de La Pampa
[4]GIGA AffectiveLab, Universidad de Zaragoza, España
{csanz, vartola, nsalazar}@lidi.info.unlp.edu.ar,
li@info.unlp.edu.ar, abrilbuffarini@gmail.com, laureanobruno@alu.ing.unlp.edu.ar, matiasmzeb@gmail.com, ,
aldanamdg@lidi.info.unlp.edu.ar, maitesegovia.a@gmail.com, astudillo@exactas.unlpam.edu.ar,
sandra@unizar.es

RESUMEN

El trabajo presenta una de las líneas de investigación, desarrollo e innovación del subproyecto "Creación de tecnologías digitales para el escenario educativo", que forma parte de uno de los proyectos acreditados del Instituto de Investigación en Informática LIDI (III-LIDI). En este subproyecto se investiga y se desarrollan aplicaciones, herramientas y metodologías que permiten integrar tecnologías digitales en los procesos de enseñar y aprender. Se detallan aquí los avances de la línea que se enfoca en la creación de experiencias educativas innovadoras con diferentes paradigmas de interacción persona-ordenador. En el último año se ha profundizado en los resultados de años anteriores, tanto en el estudio de marcos teóricos como en la aplicación y vinculación del proyecto con escenarios educativos concretos, la publicación de resultados y la formación de recursos humanos.

Palabras clave: interacción persona-ordenador, interacción tangible, realidad aumentada, realidad virtual, entornos inmersivos, inteligencia artificial, computación afectiva, escenarios educativos, juegos serios.

CONTEXTO

El subproyecto se desarrolla en el marco de un proyecto acreditado del Instituto de Investigación en Informática LIDI (III-LIDI) de la Facultad de Informática de la UNLP, titulado: "Diseño, desarrollo y evaluación de sistemas en escenarios híbridos para áreas clave de la sociedad actual: educación, ciudades inteligentes y gobernanza digital". El subproyecto que se describe aquí es el SP3: Creación de tecnologías digitales para el escenario educativo, donde se trabaja en una línea específica orientada al estudio de paradigmas de interacción persona-ordenador emergentes y sus potencialidades para crear experiencias educativas innovadoras. El subproyecto involucra otras líneas de I+D+i que articulan y aportan a la temática que aquí se presenta. Al mismo tiempo, se destaca que se cuenta con la participación de estudiantes, becarios, investigadores en formación y formados para llevar adelante las tareas propias de esta línea.

1. INTRODUCCIÓN

La interacción persona-ordenador (IPO) o también conocida como interacción humano-computadora (HCI: *Human-Computer Interaction*) es un área interdisciplinaria con una amplia gama de conceptos y conocimientos que la fundamentan. Los diseñadores de tecnologías digitales, y en particular aquellas orientadas a educación, deben considerar enfoques y metodologías de IPO que puedan atender a la cultura, necesidades y realidades del contexto para el que se diseñan (Anderson-Coto, 2024). Es por esto que la utilización de metodologías centradas en el usuario, y las técnicas de diseño participativo (un enfoque que se propone considerar los intereses, participación y valores de los miembros de la comunidad al diseñar productos) han adquirido un auge relevante para esta área. Bajo estos fundamentos en los diseños que se abordan en el subproyecto aquí presentado, se integran diferentes participantes provenientes de áreas que se vinculan con los problemas educativos a impactar.

Desde hace ya varios años en el marco del subproyecto, se analizan diferentes tecnologías y modos de interacción que posibilitan crear entornos inmersivos y aumentados para potenciar experiencias educativas innovadoras. La realidad virtual ofrece oportunidades para lograr una inmersión en un mundo 3D simulado. Cuanto más realista sea el entorno y las condiciones de integración, más se puede disfrutar de una sensación cercana a la realidad, como si lo que se viviese estuviera sucediendo realmente (Martínez-Requejo, S., Lores-Gómez, B., y Ruiz-Lázaro, 2024, citando a Lannutti, 2022). La realidad aumentada, por su parte, agrega información digital a un entorno físico. Complementa el aprovechamiento del escenario real creando experiencias combinadas. Cheng and Tsai (2013) resaltan la utilidad de la realidad aumentada tanto en el

desarrollo de habilidades espaciales y prácticas, como en la comprensión de conceptos relacionados con las ciencias. En el gradiente entre estas tecnologías se encuentran combinaciones o hibridaciones que habilitan diferentes experiencias tendiendo puentes entre el entorno físico y el digital. Algunos autores denominan a estas hibridaciones como realidades extendidas o XR (Cabero-Almenara, Valencia-Ortiz, R., y Llorente-Cejudo, 2022). La Interacción Tangible (IT) con el uso de mesas interactivas es un eje de estudio de este proyecto también (Artola, Sanz & Baldassarri, 2023; Stephanidis et al., 2019). A esto se agrega como parte de la investigación, el abordaje y análisis de qué paradigmas de interacción se adecuan mejor para el diseño de una determinada experiencia educativa de acuerdo con los objetivos, destinatarios y otros factores contextuales. Se consideran marcos teóricos, metodologías, herramientas y técnicas que posibilitan guiar la toma de decisiones y el proceso de diseño y desarrollo para la creación de tecnologías digitales para el ámbito educativo. Al mismo tiempo, se vienen estudiando y trabajando en técnicas basadas en visión por computadora, computación afectiva, inteligencia artificial, mapeo de proyección, uso de sensores, entre otras, para crear entornos inmersivos e interactivos aplicados al escenario educativo.

El desarrollo de juegos serios educativos es otro de los ejes del proyecto que se combina con el de paradigmas de interacción para ofrecer experiencias lúdico-educativas. En 2024 se trabajó en la creación de una sala de escape educativa e interactiva que es fruto de la investigación en estas temáticas. Las salas de escape educativas son experiencias lúdicas en las que grupos (entre dos y ocho personas) entran en un espacio físico o digital en el que deben buscar pistas y resolver tareas, que se relacionan con competencias a desarrollar en el aula y/o contenido didáctico. En la sección de resultados se describen más en detalle este trabajo.

Las líneas de investigación aquí presentadas guardan relación directa con las temáticas de la Maestría y Especialización en Tecnología Informática Aplicada en Educación de la misma Facultad, por lo que se desarrollan tesis y trabajos finales en estos temas. También se cuenta con una agenda de cooperación a nivel nacional e internacional, que permite profundizar el trabajo que se lleva adelante en el III-LIDI.

2. LÍNEAS DE INVESTIGACIÓN / DESARROLLO

Se mencionan aquí las principales líneas de investigación y desarrollo abordadas en el marco del proyecto:

- Paradigmas de Interacción Persona-Ordenador en escenarios educativos.
- Diseño de actividades educativas basadas en diferentes paradigmas. Las posibilidades de la Realidad Aumentada, la Interacción Tangible, la Realidad Virtual en la configuración de actividades educativas.
- La Inteligencia artificial generativa para el enriquecimiento de actividades educativas interactivas.
- Técnicas con hologramas y mapeo de proyecciones para crear entornos aumentados e interactivos.
- Juegos serios como actividades educativas. Multisensorialidad a través de juegos y actividades educativas basadas en diferentes paradigmas de interacción.
- Juegos móviles usando aprendizaje situado y basado en la experiencia.
- *Exergames* para promover la actividad física.
- Diseño de Objetos Activos para IT. Tipos de *feedback* en objetos activos. Uso de actuadores y sensores.
- Entornos inmersivos e interactivos basados en realidad virtual y aumentada. *Embodied Interaction.*
- Sistemas educativos adaptativos y sistemas recomendadores para recursos educativos.
- Computación afectiva en entornos digitales para el escenario educativo.

Estas líneas se basan en el trabajo conjunto de docentes investigadores, becarios, tesistas, pasantes y estudiantes que participan. Al mismo, tiempo se aborda en muchos de estos proyectos un diseño participativo con docentes y estudiantes de diferentes niveles educativos que intervienen con propuestas y en el diseño.

3. RESULTADOS OBTENIDOS

En el año 2024 se ha logrado avanzar con diferentes resultados relevantes del proyecto. Se articuló con un proyecto de extensión de la Facultad de Informática en el que se aplican y se llevan al campo juegos serios y experiencias creadas en el marco de las líneas de investigación mencionadas.

Se listan a continuación varios de los resultados alcanzados.

3.1 Sala de Escape Educativa

En julio de 2024 se inaugura una sala de escape en el Centro de innovación y Transferencia Tecnológica (CIyTT)[1] en uno de los espacios del III-LIDI. El objetivo educativo se centra en los temas de alimentación saludable y nutrición. La sala incluye una serie de desafíos para resolver en un tiempo máximo de 15 minutos. Se puede jugar con grupos de hasta 4 personas (Figura 1 - Izquierda). Involucra diferentes tecnologías digitales creadas en el proyecto:

- Un panel interactivo con información relevante sobre grupos de alimentos y sus nutrientes, la gráfica de alimentación saludable y la presentación de la Ley de etiquetado frontal,

[1] https:/ciytt.info.unlp.edu.ar/

con sus sellos y significado (Sanz, Artola & Bruno, 2023) (Figura 1 - Derecha).

- Una aplicación de Realidad Aumentada que combina el rostro de uno de los jugadores con un aumento que da cuenta del nivel de desempeño del jugador al finalizar la sala.
- Un controlador de tiempo automático durante el juego.

Además, la sala de escape se utilizó para un experimento vinculado a las investigaciones del área de computación afectiva. Se propone detectar estrés en las personas que están jugando. El experimento se diseñó para ser aplicado en sesiones específicas de las sala de escape, con un momento inicial de explicación en la que los jugadores se colocan sensores de pecho de ritmo cardíaco, y una entrevista posterior a la sala de escape. Los resultados de este experimento están siendo analizados para su posterior difusión. Se realizó como parte de una práctica profesional supervisada de un estudiante de Ingeniería en Computación (Zeballos, 2024).



Figura 1 – Izquierda: escenografía de la sala de escape. Derecha: usuarios usando el panel interactivo mientras juegan

3.2 Reflector mágico

Se trata de una iniciativa que aborda el trabajo con sensores y se enfoca en el paradigma de interacción tangible y mapeo de proyecciones. Se utiliza un dispositivo creado ad-hoc que simula ser un reflector y posee un sensor acelerómetro y giroscopio (Figura 2). A partir de su movimiento se determina el área donde se hará la proyección con un proyector convencional oculto en el entorno. Así, cuando el usuario utiliza el reflector mágico y lo apunta hacia un lugar concreto del espacio físico, la proyección con el proyector oculto, se realiza para mostrar narrativas digitales sobre el espacio físico. El proyecto es escalable para generar entornos inmersivos e interactivos con diferentes narrativas digitales y educativas. Se realizó en 2024 un trabajo para mostrar y presentar los personajes del libro "El principito". Para ello se diseñó un panel físico con diferentes planetas visitados por este personaje, incluyendo una adaptación con un nuevo planeta (Figura 3). Cuando el reflector enfoca cada planeta, aparece su personaje principal, y se da cuenta de cómo el Principito descubre el mundo de los adultos a través de su interacción con estos. Se impacta en el área de educación ya que ofrece una forma diferente de trabajar con narrativas digitales. En el proyecto participaron docentes y estudiantes. Se participó en la Noche de los Museos junto con el Museo de Física de la UNLP, y también en la Expo Ciencia y Tecnología de la Facultad de Informática.

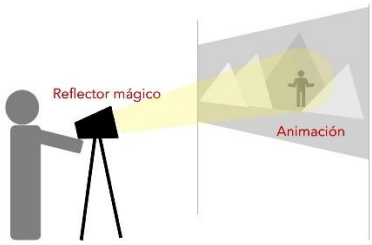


Figura 2 – Esquema de la propuesta de reflector mágico



Figura 3 – Muestra el panel con los planetas y una persona utilizando el reflector mágico

3.3 Proyectos interactivos con estudiantes de escuela media

Durante el 2024 se llevó adelante un proyecto llamado "Semillero de innovadores" en el contexto de las actividades del CIyTT, destinado a estudiantes de escuela media (Sanz et al., 2024). Se trabajó con un grupo reducido de estudiantes ya que se trató de una prueba piloto. La experiencia se realizó de abril a julio y se enfocó en el desarrollo de competencias del área de programación y electrónica, a través de una propuesta proyectual del tipo "aprender haciendo". Se llevaron adelante dos proyectos: a) con los estudiantes avanzados, una casa maquetada inteligente, con una *app* móvil para controlar las luces *leds*, la apertura de la puerta, el timbre, etc. (Figura 4 - Izquierda); y b) con los estudiantes más jóvenes, el diseño e implementación de interfaces de papel con el simulador llamado *tinkercad*. Los proyectos fueron presentados por los estudiantes en la Jornada de innovación del CIyTT. Las Figuras 4 Izquierda y Derecha muestran los proyectos y los estudiantes presentando.

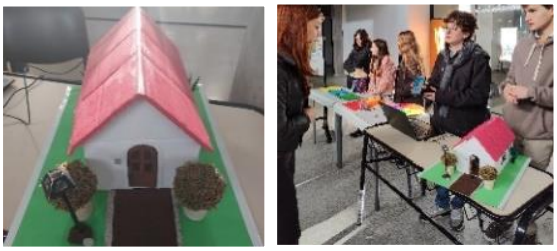


Figura 4 – Izquierda: Maqueta Casa inteligente; Derecha: Estudiantes participantes presentando

3.4 Trabajo con escuelas con los proyectos creados

En el 2023 se había creado un ecosistema de herramientas orientado a fomentar la alimentación

saludable, temas de nutrición, y la importancia de la actividad física en la vida de las personas. En 2024 se desarrollaron talleres con el uso de una mascota interactiva, "Indy Bob" que fue creada en el contexto de una convocatoria de proyectos de innovación con alumnos de la facultad 2023 (Sanz, Artola & Baldassarri, 2024). En el marco de talleres con escuelas, también se utilizó un juego sobre una mesa interactiva con dos desafíos. El primer desafío consiste en colocar los sellos de la Ley de etiquetado frontal a una serie de alimentos que aparecen sobre la superficie de la mesa (interfaz). El segundo desafío consiste en una guerra de alimentos. Se trata de una competencia, en la que alternadamente cada jugador dispara alimentos mediante un cañón que aparece en la interfaz (superficie de la mesa). El otro jugador debe atrapar con una canasta física solo aquellos alimentos que correspondan a un grupo específico; en caso atrapar un alimento de otro grupo, se pierden puntos. En 2024 se llevaron adelante numerosas actividades que articularon con un proyecto de extensión de la Facultad de Informática que aborda esta temáticas. Se realizaron 3 visitas para trabajar con alumnos de primer grado de una escuela con la mascota interactiva y nutricionistas, se llevó adelante un trabajo con el colegio Nacional de La Plata y jóvenes del programa Envión de la provincia de Buenos Aires, entre otras vinculaciones que permitieron la integración de resultados de este proyecto.

Además, en función del trabajo con una de las escuelas de la UNLP se extendió el proyecto Murales orientado a la creación de murales sobre una mesa interactiva. Esta nueva versión incorpora nuevos objetos y mejoras en su funcionamiento (Segovia, Artola y Sanz, 2024).

3.5 Avance en las investigaciones

Al mismo tiempo se continúo avanzando en temas de investigación que se relacionan con las líneas aquí mencionadas. Se finalizó una tesis doctoral sobre recomendación de piezas musicales basadas en etiquetado emocional (Ospitia-Medina, 2024). Se entregó una tesis doctoral relacionada con gamificación en entornos virtuales que cuenta con una dirección compartida con profesores de la Universidad de Zaragoza y la UNLP. Actualmente, se está esperando su evaluación (Glenda Vera et al., 2024). Se continúa con una tesis doctoral enfocada en la detección de la emoción de aburrimiento mientras se visualizan videos, en particular educativos (Astudillo et al., 2023). Este último trabajo se enmarca en los temas de computación afectiva y es una tesis de cotutela con la Universidad de Zaragoza. Se está desarrollando una tesis de maestría en temas de realidad virtual (Lanata, Rucci & Sanz, 2024), que se vincula con el programa de extensión Huellas patrimoniales[2] (en el que se participa). En dicho programa se han desarrollado tres juegos de RV llamado HuVi Tango y Casa Curutchet, HuVi Ischigualasto y Talampaya y HuVi Parque Nacional Iguazú (Sanz et al., 2023).

Una de las investigadoras en formación del proyecto está abordando el estudio de detección de movimientos y gestos de la mano para el manejo de la cámara en juegos en primera persona, del tipo exergames (Del Gener, Sanz & Iglesias, 2023). Otra de las investigadoras está finalizando su tesis de maestría en temas de realidad aumentada (Salazar Mesía & Sanz, 2023). Se publicaron los resultados del juego EmpoderAR a partir de su uso con estudiantes de escuela secundaria (Sanz, Artola & Ibañez, 2024).

Cabe aclararse que todos estos resultados se desarrollaron en vinculación con las líneas aquí presentadas y los juegos y experiencias están disponibles para quienes participan de las actividades del CIyTT.

En cuanto a los proyectos vinculados y los acuerdos de cooperación, el III- LIDI participa en los siguientes:

- Se cuenta con un acuerdo de colaboración con la Universidad de Zaragoza y la Universidad de Islas Baleares.
- Se ha obtenido financiamiento en el marco de una convocatoria K171 de Erasmus+ para movilidades entre la Universidad de Zaragoza y la Universidad Nacional de La Plata en 2024. Se ha presentado un nuevo pedido de financiamiento de movilidades a través de un K171 en el marco del programa UNITA-GEMINAE del que se participa en 2025.
- La Dra. Sanz es miembro colaborador del grupo de Investigación en Interfaces Avanzadas (AffectiveLab).
- Se asesora en el proyecto de la Universidad de Costa Rica: "Juegos serios basados en interfaces tangibles con objetos activos". Cod-510-C2-321
- Se participa en la RedAUTI: Red temática en Aplicaciones y Usabilidad de la Televisión Digital Interactiva.
- Se participa del proyecto TEMOR: Ecosistema tecnológico para el reconocimiento del estado de ánimo de los pacientes durante el proceso de rehabilitación cardíaca. Proyecto de la Universidad de Zaragoza, aprobado en 2022. Duración 2 años.
- Se participa del proyecto: Experiencias Lúdicas con Agentes Sociales Interactivos y Robots: Aprendizaje Social y Comunicación Intergeneracional (PLEISAR). Dirigido por la Dra. Cerezo de la Universidad de Zaragoza.
- Se cuenta con dos acuerdos de cooperación con: 1- la Facultad de Arquitectura para la que se realizaron dos proyectos culminados en 2024: uno basado en mesa interactiva con sensores y proyecciones, y el segundo con mapeo de proyecciones; 2-con el Museo de Ciencias Naturales de La Plata para quienes se está realizando una instalación interactiva basada en mapeo de proyecciones vinculada también a temas de interés de las redes de UNITA-GEMINAE.

4. FORMACION DE RRHH

[2] https://www.huellaspatrimoniales.econo.unlp.edu.ar/

Se trabaja en la formación de recursos humanos a través de la dirección de tesis de doctorado, maestría, trabajos de especialización y tesinas de grado; también con becas de diferentes organismos de ciencia y técnica. En 2024, se han completado
tesis doctorales, una de forma completa y la otra esperando su evaluación. Se entregó un trabajo de Especialización en TIAE que está esperando evaluación (Gubaro, Sanz & Artola, 2024). Se han concluido dos prácticas profesionales supervisadas de Ingeniería en Computación en temas relacionados al proyecto y se está desarrollando otra de Analista en TIC.

5. BIBLIOGRAFÍA

- Anderson-Coto, M.J., Salazar, J., López, J.L.S., Sedas, R.M., Campos, F., Bustamante, A.S., Ahn, J. (2024). Towards culturally sustaining design: Centering community's voices for learning through participatory design. International Journal of Child-Computer Interaction 39, 100621. https://doi.org/10.1016/j.ijcci.2023.100621
- Archuby, F., Sanz, C., & Manresa-Yee, C. (2023). DIJS: Methodology for the Design and Development of Digital Educational Serious Games in IEEE Transactions on Games, doi: 10.1109/TG.2022.3217737. Early access.
- Artola, V., Sanz, C., & Baldassarri, S. (2023). "A Novel Tangible Interaction Authoring Tool for Creating Educational Activities: Analysis of Its Acceptance by Educators," in IEEE Trans. on Learning Technologies, vol. 16, no. 5, pp. 633-647, doi: 10.1109/TLT.2022.3216117
- Astudillo, G. J., Asiain, D., Sanz, C. V., Ponce, J., Beltrán, J. R., & Baldassarri, S. (2023). Capturing physiological data for detecting boredom in educational contexts. XXIII Int. Conference on HCI, Lleida, Spain. ACM Digital Library. DOI: 10.1145/3612783.3612811.
- Cabero-Almenara, J., Valencia-Ortiz, R., & Llorente-Cejudo, C. (2022). Ecosistema de tecnologías emergentes: realidad aumentada, virtual y mixta. Revista Tecnología, Ciencia Y Educación, (23), 7–22. https://doi.org/10.51302/tce.2022.1148
- Centro de Innovación y Transferencia Tecnológica, FI, UNLP, https:/ciytt.info.unlp.edu.ar/
- Cheng, K.-H., & Tsai, C.-C. (2013). Affordances of augmented reality in science learning: Suggestions for future research. J. of Science Education and Technology, 22(4), 449–462. https://doi.org/10.1007/s10956-012-9405-9
- Del Gener A., Sanz C., & Iglesias L. (2023). Propuesta de un Gamepad para Sensar Movimientos del Jugador y su Integración a un Exergame. Revista AIPO, 4(2), 68-77.
- Gubaro, M., Sanz, C. V., & Artola, V. (2024). Juegos Serios Educativos Digitales Orientados a la Enseñanza de la Matemática en Educación Secundaria. In XXX Congreso Argentino de Ciencias de la Computación (CACIC 2024).
- Lanata, L., Sanz, C. & Rucci, A. (2024). Un viaje a la era del hielo: un recorrido virtual educativo por el Parque Nacional los Glaciares para alumnos del ciclo básico del nivel secundario. Prop. de tesis de maestría.
- Martínez-Requejo, S., Lores-Gómez, B., & Ruiz-Lázaro, J. (2024). Efectividad de las tecnologías inmersivas para potenciar el aprendizaje en educación superior: una revisión sistemática. Edutec, Revista Electrónica y Tecnología Educativa, (90), 54-73. https://doi.org/10.21556/edutec.2024.90.3391
- Ospitia Medina, Y. (2024). Desarrollo de un modelo de elicitación de emociones a partir de las características de la música (Tesis Doctoral, UNLP). https://doi.org/10.35537/10915/164064
- Salazar Mesía, N. A., & Sanz, C. V. (2023). Avances en la herramienta de autor AuthorAR para la creación de actividades educativas con geolocalización. In XVIII Congreso Nacional de Tecnología en Educación y Educación en Tecnología-TE&ET 2023.
- Sanz, C. V., Rucci, A. C., Mazza, M. A., Renzella, L. B., Romero, A., Ravea, N. M. S., & Artola, V. (2023). Realidad virtual para acercar el patrimonio argentino: el caso de HuVi Tango y Casa Curutchet. In XVIII Congreso TE&ET 2023.
- Sanz, C., Artola, V. y Bruno, L. (2023). Seminario III-LIDI presentación de "Panel interactivo para abordar temas de alimentación saludable y nutrición". Fecha: 04-08-2023.Disponible en: https://shorturl.at/AWucK
- Sanz, C., Artola, V., & Ibañez, B. (2024) EmpoderAR un Juego Serio de Realidad Aumentada para Acercar a Jóvenes a la Ciencia, Interacción Rev. digital de AIPO, 5(1), 6-13.
- Segovia M., Artola V. & Sanz C. (2024) Informe técnico del proyecto Murales.
- Lannutti, E. D. (2022). Robótica: De la ciencia ficción a la realidad científica (Vol. 13). EDIUNC.
- Sanz, C. V., Artola, V., & Baldassarri, S. (2024). Developing educational serious games as a proposal for university training in the area of Child Computer Interaction. International Journal of Child-Computer Interaction, 41, 100672.
- Sanz, C., Artola, V., Medina, S., Zeballos, M., & Lombardo, S. (2024). Crear con tecnologías digitales para el desarrollo del pensamiento computacional. El caso del proyecto de semillero de innovadores.
- Stephanidis, C., Salvendy, G., Antona, M., Chen, J. Y., Dong, J., Duffy, V. G., et al. (2019) Seven HCI grand challenges. International Journal of Human–Computer Interaction, 35(14), 1229–1269 (2019).
- Vera-Mora, G., Sanz, C. V., Coma-Roselló, T., & Baldassarri, S. (2024). Model for Designing Gamified Experiences Mediated by a Virtual Teaching and Learning Environment. Ed. Sciences, 14(8), 907. https://doi.org/10.3390/educsci14080907
- Zeballos, M., & Sanz, C. (2024). Informe final Práctica Profesional Supervisada: "Prototipo funcional de un firmware de registro de sensores fisiológicos para el análisis de emociones. El caso de una sala de escape inmersiva y educativa.

Deserción en las carreras de Ingeniería: determinación de perfiles de los estudiantes. Estrategias para disminuir la deserción.

Minnaard Claudia, Comoglio Marta

Instituto de Investigaciones de Tecnología y Educación (IITE)

Facultad de Ingeniería – Universidad Nacional de Lomas de Zamora

Buenos Aires, Argentina

RESUMEN

En esta oportunidad, el proyecto de investigación se realiza en forma conjunta con la Facultad de Ciencias Agrarias de la Universidad Nacional de Lomas de Zamora, ya que existen indicios suficientes que avalarían la hipótesis de que las causas de la deserción de los estudiantes son similares. A partir de la aplicación de métodos prospectivos (análisis estructural y método de los escenarios) se definen las variables claves, así como posibles escenarios futuros con su respectiva probabilidad.

Palabras clave: Deserción; Métodos Prospectivos

CONTEXTO

En la actualidad la deserción estudiantil es uno de los problemas que aqueja a la mayoría de las instituciones de educación superior en toda Latinoamérica. El número de investigaciones que a la fecha abordan esta problemática es grande, y dan cuenta del número de estudiantes que no logran culminar sus estudios universitarios y de los costos sociales relacionados a este fenómeno.

La asimetría entre el grupo de graduados con la cantidad de ingresantes se ve además reforzada por otro factor, que es el alargamiento de las carreras. Tal como surge de algunas investigaciones (Pagura, J. [et al] 2000: 129-141 ; Di Gresia, L.[et al] 2002: 20-32; Di Gresia L [et al], 2003: 3; Giovagnolli, P (2002): 17-21; Porto, A. [et al] (2004) : 12; Cortes et al (2019); Gallegos et al (2018); Seminara et al(2018)) dentro del grupo que llega a graduarse, existe un importante porcentaje de alumnos que prolonga su permanencia en la universidad más allá de los tiempos establecidos por los planes de estudio.

Dos de los indicadores usualmente utilizados para medir la eficiencia de la enseñanza en las universidades nacionales - "egresados en función de ingresantes y "duración teórica – real de carreras" - también vienen generando en la región, preocupación en virtud de los bajos niveles de rendimiento que reflejan.

En relación a la deserción, son varias las perspectivas teóricas que intentan explicar sus causas, atribuyéndola a diferentes factores.

Ambos aspectos – abandono y rendimiento - constituyen un asunto complejo que ha merecido tratamiento en varios trabajos de investigación.

Algunos autores (Ewell, 1995: 12; García de Fanelli, 2003), aseguran que para arrojar luz sobre esta complejidad se requieren dos momentos:

a) Análisis: Comprende la descripción de los fenómenos (abandono y rendimiento académico)

b) Explicación: De los factores que actúan en su determinación García de Fanelli (2004: 71), afirma que en primer lugar, se debe distinguir entre el abandono del sistema (Educación Superior) del que realiza el estudiante de una cierta organización educativa en particular. En este último caso, se está ante la situación de

cambio de universidad, lo que en términos de los registros institucionales aparecerá como una baja académica, pero en términos del sistema constituye simplemente una transferencia entre instituciones. Según la autora este tipo de abandono es muy frecuente por motivos vocacionales, financieros o por insatisfacción frente a la oferta educativa.

Considera que, si se tiene en cuenta la gestión de la política académica de la institución universitaria, lo que fundamentalmente interesa a sus autoridades es el plano institucional, en tanto que cada abandono impacta negativamente sobre los indicadores institucionales de calidad y financieros.

En segundo lugar, la autora afirma:

"*...otra forma que asume la interrupción de los estudios por parte de los alumnos, es cuando el estudiante deja de cursar o rendir por un período para volver a retomar los estudios tiempo después. Estas interrupciones provisorias pueden dar lugar a que transcurrido un período de tiempo el estudiante retorne a la misma institución o bien reinicie sus estudios en otra. En este último caso, se constituirá también en un abandono definitivo para la universidad, pero provisorio para el sistema*" (García de Fanelli, 2004: 73).

En este sentido, García de Fanelli considera que existen tres momentos en los que se puede situar el abandono:

1) Cuando el estudiante manifiesta deseos de continuar estudios de Educación Superior, inscribiéndose en la universidad o en su curso de admisión, pero no llega nunca a cursar o a aprobar alguna materia a lo largo del primer año;

2) Cuando el estudiante ha cursado y aprobado algunas materias a lo largo del primer año pero luego abandona y

3) El abandono de un estudiante a partir del segundo año.

En relación al análisis del ritmo en el progreso de la carrera se pueden utilizar los siguientes indicadores de rendimiento académico (García de Fanelli, 2004: 74)

a) Promedio anual de asignaturas rendidas desde el momento de ingreso

b) Coeficiente entre las asignaturas aprobadas y las rendidas

c) El promedio de calificaciones

d) La duración real de la carrera respecto de la teórica.

La propuesta de García de Fanelli es que luego de describir ambos fenómenos - abandono y rendimiento académico-, se dé comienzo a una etapa de análisis explicativo cuyos resultados contribuyan al diseño de estrategias organizacionales que mejoren la eficiencia universitaria.

Por otra parte, Cortés-Cáceres, S., Álvarez, P., Llanos, M., & Castillo, L. (2019) consideran que la deserción o abandono se puede explicar a través de cinco dimensiones la psicológica, sociológica, económica, institucional y de interacción.

- La dimensión psicológica: establece que los rasgos de personalidad del estudiante diferencias aquellos que completan sus estudios de aquellos que no, entre estos rasgos, están las intenciones, las normas subjetivas, las conductas de logro, el autoconcepto académico, la percepción de dificultad de los estudios y el nivel de aspiraciones.

- La dimensión sociológica: enfatiza la influencia que tienen factores externos al individuo sobre la retención, los que se suman a los psicológicos mencionados de la dimensión anterior, los aspectos relevantes son la integración social, el medio familiar, la inserción en la universidad, la congruencia normativa, el compromiso institucional y el apoyo de pares.

- La dimensión económica: guarda relación por un lado con la capacidad o incapacidad del estudiante de pagar la carrera universitaria, y con la obtención de subsidios de arancel, y por otro, el costo de oportunidad como un beneficio derivado de la educación universitaria.

- La dimensión institucional: se refleja en las características de la institución universitaria, los servicios estudiantiles que ofrece, los indicadores de docencia, enseñanza y calidad, la infraestructura y las experiencias de los estudiantes en el aula.

- La dimensión de interacción: reconoce al abandono como un resultado de una serie de interacciones con los sistemas académicos y sociales de la universidad.

1. INTRODUCCIÓN

El Sistema Argentino de Educación Universitaria está conformado actualmente por 131 instituciones educativas. (Secretaría de Políticas Universitarias SPU – 2019)

De acuerdo a los datos proporcionados por la SPU, la evolución de la población estudiantil universitaria (Período 2001- 2017) se observa en la Figura 1. La regresión lineal planteada muestra un incremento promedio de 43376 nuevos estudiantes por año

.

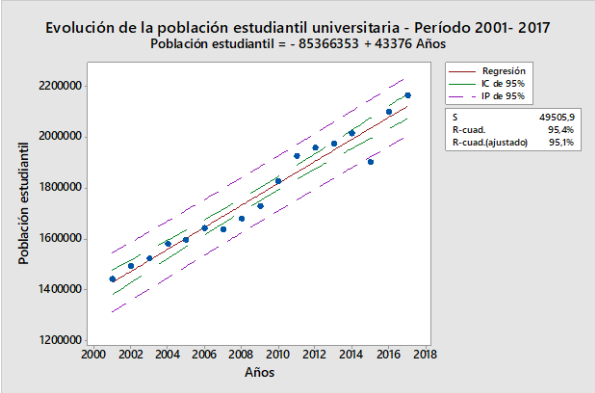


Figura 1: Evolución de la población estudiantil universitaria. Fuente: Elaboración propia sobre datos SPU

En este marco, Celada (2020) considera que *"basada en los datos brindados por la Secretaría de Políticas Universitarias (SPU, 2011), la tasa de deserción en el ámbito universitario estatal es de aproximadamente un 80 %; mientras que la tasa de deserción universitaria en el ámbito privado es aproximadamente del 60 %"*.

Cabe destacar que Garcia Fanelli et al (2015) considera que "*Argentina se destaca en América latina por tener una de las tasas brutas de educación superior más altas, similar a la de los países desarrollados, fruto de un sistema no selectivo de admisión. No obstante, la igualdad de oportunidades en el acceso no ha sido suficiente para garantizar la equidad en los resultados"*.

Estudiantes, nuevas/os inscriptas/os y egresadas/os por nivel y sector de gestión. Año 2020

		Estudiantes	Nuevas/os Inscriptas/os	Egresadas/os
Total general ›	Total	2.476.945	681.804	137.525
Pregrado y grado ›	Total	2.318.255	641.929	122.679
	Estatales	1.872.591	513.151	73.832
	Privadas	445.664	128.778	48.847
Posgrado ›	Total	158.690	39.875	14.846
	Estatales	114.973	24.600	8.275
	Privadas	43.717	15.275	6.571

Figura 2: Población estudiantil universitaria clasificados como Estudiantes, Nuevos inscriptos y Egresos. Fuente: Departamento de Información Universitaria – DNPeIU - SPU

De acuerdo a los datos proporcionados por el Departamento de Información Universitaria (DNPeIU – SPU) de los nuevos inscriptos a carreras de pregrado y grado, el 23,9% opta por carreras vinculadas a la ciencia o la tecnología. De los nuevos inscriptos a carreras de pregrado y grado en 2019, el 61,9% continúan sus estudios en 2020. Asimismo, el 22,2% de los nuevos inscriptos de grado opta por otra oferta académica uno o dos años después de su ingreso a una determinada carrera. Solamente el 25,1% de los egresados de grado lo hace en el tiempo teórico esperado para una determinada carrera.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

En el presente proyecto se plantean los siguientes objetivos así como los métodos para llevarlos a cabo:

- Indagar y evaluar las percepciones de los estudiantes con respecto a las causas de abandono en las carreras de ingeniería.
- Identificar factores clave que favorecen la deserción en los distintos tramos de la carrera.
- Desarrollar y validar una herramienta que permita diseñar y evaluar propuestas pedagógicas en carreras de ingeniería que favorezcan la retención de los estudiantes.
- Diseñar objetos de aprendizaje que contribuyan al desarrollo y articulación de competencias de ingreso desde el nivel medio a las carreras científicas tecnológica, aplicarlos al proceso de enseñanza y evaluar el impacto en la percepción de los alumnos y en los aprendizajes.
- Establecer las variables clave y ponderar escenarios para diseñar estrategias para disminuir la deserción.

3. RESULTADOS OBTENIDOS Y ESPERADOS

Se realiza un estudio exploratorio y descriptivo de aproximación al conocimiento de las ideas del alumnado mediante técnicas de análisis cualitativo. Este tipo de investigaciones se propone describir el comportamiento de variables y/o identificar tipos o pautas características resultantes de las combinaciones de un cierto número de ellas. Las investigaciones descriptivas se ocupan entonces de identificar las variables relevantes del objeto o asunto investigado, y luego de averiguar cómo se comportan dichas variables. (Ynoub, 2007; Mombrú, 2013)

En este último sentido, se utilizan preguntas abiertas categorizadas en dimensiones sobre las cuales se debaten posibles implicancias para la permanencia estudiantil.

Asimismo, se incluyen preguntas asociativas en las cuales los estudiantes pueden mencionar factores que asocian con el bienestar y con el abandono.

En una etapa posterior, se busca determinar no sólo el comportamiento de las variables sino además la dependencia o vinculación que unas variables mantienen con otras.

De manera más específica, interesa probar si el comportamiento de una variable (o varias) puede funcionar como factor o causa explicativa del comportamiento de otra (u otras) variable.

Estudio prospectivo: A través del método de análisis estructural (matriz de Impactos Cruzados, Multiplicación aplicada a una Clasificación) y del Método de Escenarios se proyectará a través de paneles de expertos un modelo que identifique variables claves del sistema y el escenario posible para disminuir la deserción y aumentar la retención. Se utilizan las herramientas MICMAC y SMIC-PROB-EXPERT desarrollados por Lipsor.

4. FORMACIÓN DE RECURSOS HUMANOS

Se prevé la formación de un becario EVC – CIN.

5. BIBLIOGRAFÍA

- CELADA, V. L. (2020). Acerca de las causas de deserción universitaria en Argentina a principios del siglo XXI, de las políticas implementadas y nuevas propuestas de retención de población estudiantil.

- CORTÉS - CÁCERES, S., ÁLVAREZ, P., LLANOS, M., & CASTILLO, L. (2019). Deserción universitaria: La epidemia que aqueja a los sistemas de educación superior. Revista perspectiva, 20(1), 13-25.

- DI GRESIA, L. y PORTO, A. (2003): Dinámica del desempeño Académico. Universidad Nacional de La Plata. Documento de trabajo Nº 49. [en línea] <http://www.aaep.org.ar/espa/anales/resumen

04/04/Porto- DiGresia-Armengol.pdf > [consulta: 12 de septiembre de 2022]

- DI GRESIA, L, PORTO, A. y RIPANI, L. (2002): Rendimiento de los estudiantes de las universidades públicas argentinas. Universidad Nacional de la Plata [en línea] noviembre, 2002. < http:// www.depeco.econo.unlp.edu.ar/doctrab/doc45.pdf > [consulta: septiembre, 10, 2022]

- GALLEGOS, J. A., CAMPOS, N. A., CANALES, K. A., & GONZALEZ, E. N. (2018). Factores determinantes en la deserción universitaria. Caso Facultad de Ciencias Económicas y Administrativas de la Universidad Católica de la Santísima Concepción (Chile). Formación universitaria, 11(3), 11-18.

- GARCIA DE FANELLI, A. (2000) Los indicadores en las políticas de reforma universitaria argentina: balance de la situación actual y perspectivas futuras. En KISILEVSKY, M. (coord.) Indicadores Universitarios. Tendencias y experiencias internacionales, pp. 17-54, Buenos Aires: EUDEBA.

- GARCIA DE FANELLI, A.M. (2003): El abandono y rendimiento académico como fenómenos complejos. En 2do. Taller 45º Aniversario de las Universidades Privadas CRUP Impacto de las Universidades Privadas en la sociedad Argentina, agosto, 2003 Mendoza: Universidad de Champagnat.

- GARCIA DE FANELLI, A.M., (2004): Indicadores y estrategias en relación con el abandono y la graduación universitarias. En: MARQUIS C. (2004): La Agenda Universitaria: Propuestas de políticas públicas para la Argentina. Buenos Aires: Universidad de Palermo Colección Educación Superior, pp. 65-90

- GARCIA DE FANELLI, A. M. (2005). Universidad, Organización e Incentivos: Desafíos de la política de financiamiento frente a la complejidad institucional. Buenos Aires: Miño y Dávila-F/OSDE pp. 32-57.

- GIOVAGNOLI, P. (2002): Determinantes de la deserción y graduación universitaria: una aplicación utilizando modelos de duración. Documentos de trabajo.

-LIPSOR. Métodos de prospectiva estratégica. En: http://es.laprospective.fr/Metodos-de-prospectiva.html

- MOMBRÚ, A. & MARGETIC, A. (2013) El hacedor de tesis. Avellaneda, L.J.C. Ediciones

- PAGURA J. A; QUAGLINO M. e ITURBIDE, D. (2000): Un modelo estadístico para evaluar tiempos medios empleados en culminar etapas en la universidad. Instituto Rosario de Investigaciones en Ciencias de la Educación. Rosario. Revista IRICE nº 14 pp.129-141.

- PASCAL, O. (2010). Las Tecnologías de la Información y Comunicación (TIC) y su aplicación a la enseñanza técnica a través de modelos de enseñanza centrados en el alumno (Doctoral dissertation, Universitat Politècnica de València).

- SEMINARA, M. P., & APARICIO, M. T. (2018). Deserción Universitaria ¿Un concepto equivoco?: Revisión de estudios latinoamericanos sobre conceptos alternativos.

- SEMINARA, M. P. (2021). De los efectos de la pandemia COVID -19 sobre la deserción universitaria: desgaste docente y bienestar psicológico estudiantil. Revista Educación Superior Y Sociedad (ESS), 33(2), 402-421. https://doi.org/10.54674/ess.v33i2.360

- Síntesis de Información Estadística Universitaria 2020-2021. En https://www.argentina.gob.ar/sites/default/files/sintesis_2020-2021_sistema_universitario_argentino.pdf

- YNOUB, R. (2007) El Proyecto y la metodología de la investigación. Buenos Aires, Ed. Cengage Learning. Cap. VI.

Diseño de objetos de aprendizaje usando herramientas de Inteligencia artificial.

Minnaard Claudia[1], Torres Zulma[1], Minnaard Vivian[2]

[1]Instituto de Investigaciones de Tecnología y Educación (IITE)

Facultad de Ingeniería – Universidad Nacional de Lomas de Zamora

[2]Universidad FASTA

Buenos Aires, Argentina

minnaardclaudia@gmail.com , iite@ingenieria.unlz.edu.ar

RESUMEN

El proyecto de investigación se centra en el desarrollo de objetos de aprendizaje adecuados para las carreras de ingeniería utilizando herramientas de Inteligencia Artificial. Una vez diseñados y aplicados se evalúan los resultados obtenidos.

Palabras claves: Inteligencia artificial, Objetos de aprendizaje, Ambientes virtuales de aprendizaje.

CONTEXTO

El proyecto se desarrolla en el Instituto de Investigaciones de Tecnología y Educación de la Facultad de Ingeniería de Universidad Nacional de Lomas de Zamora. Desde la fecha de su creación el Instituto, en 2009, resulta ser un ámbito de investigación y desarrollos concretos sobre ambientes virtuales y exploración de aplicaciones de software para la enseñanza en carreras tecnológicas, en particular la ingeniería.

Son diversos los autores que consideran que "desde noviembre de 2022, con la aparición de la herramienta ChatGPT, hubo un aumento exponencial en el uso de la inteligencia artificial en todos los ámbitos. Aunque ChatGPT es solo una de las muchas tecnologías generativas de inteligencia artificial, su impacto en los procesos de enseñanza y aprendizaje ha sido notable" (García Peñalvo, (2024); González, C. S. (2004); Heredia, M. O. T., Correa, Y. K. D., & Carballo, E. V. (2023)).

Para García Peñalvo (2024) " la IA generativa tiene como objetivo la generación de contenidos. Los modelos de lenguaje que se usan para este fin se entrenan para determinar qué elementos tienen una mayor probabilidad de aparecer cerca de otros. Para generar sus respuestas, evalúan grandes corpus de datos, lo que les permite satisfacer a las solicitudes con respuestas que entran dentro de una probabilidad determinada para el corpus del entrenamiento, es decir, sin implicar un razonamiento, de forma que, aunque la respuesta sea coherente, no implica que sea siempre correcta. Esta característica debe tenerse en cuenta en cualquier ámbito, pero especialmente en los usos educativos de estas herramientas".

El Consenso de Beijing sobre la inteligencia artificial y la educación (UNESCO, 2019) destaca 44 recomendaciones, agrupadas en distintos aspectos que pueden ayudar a entender la magnitud de la tarea:

- planificación de la IA en las políticas educativas;
- la IA para la gestión y la impartición de la educación;
- la IA para apoyar la docencia y a los docentes;
- la IA para el aprendizaje y la evaluación del aprendizaje;
- el desarrollo de valores y competencias para la vida y el trabajo en la era de la IA;
- la IA para ofrecer oportunidades de aprendizaje permanente para todos;

- promoción del uso equitativo e inclusivo de la IA en la educación;

- IA con equidad de género e IA para la igualdad de género; y velar por el uso ético, transparente y verificable de los datos y algoritmos educativos

La UNESCO exige intrínsecamente "un enfoque de la IA centrado en el ser humano. Su objetivo es incluir el papel desempeñado por la IA en la solución de las desigualdades actuales en materia de acceso al conocimiento, la investigación y la diversidad de las expresiones culturales, y garantizar que la IA no amplíe la brecha tecnológica dentro de los países y entre ellos. La promesa de la "IA para todos" debe permitir que cada cual pueda sacar provecho de la revolución tecnológica en curso y acceder a sus frutos, fundamentalmente en términos de innovaciones y conocimientos". (Consenso de Beijing, 2019)

En la Guía para el uso de IA generativa en educación e investigación (UNESCO, 2023) se destaca que "la inteligencia artificial generativa (IAGen) irrumpió en la escena pública a finales de 2022 con el lanzamiento de ChatGPT, que se convirtió en la aplicación de más rápido crecimiento en la historia. Con el poder de emular las capacidades humanas para generar resultados en formatos como texto, imágenes, videos, música y códigos de software, estas aplicaciones de IAGen han causado sensación".

Las tecnologías específicas en las que se basa la IAGen forman parte de la familia de tecnologías de IA denominada aprendizaje automático (AA), que utiliza algoritmos que le permiten mejorar de forma continua y automática su rendimiento a partir de los datos. El tipo de AA que ha dado lugar a muchos de los avances en IA que hemos visto en los últimos años, como la IA de reconocimiento facial, se conoce como redes neuronales artificiales (RNAs), que se inspiran en el funcionamiento del cerebro humano y sus conexiones sinápticas entre neuronas.

1. INTRODUCCIÓN

En la actualidad, el auge de las Inteligencias Artificiales (IA) ha implicado un cambio significativo en las metodologías de enseñanza dentro del ámbito de la educación universitaria. El objetivo principal de esta transformación es explorar y comprender el desarrollo y la implementación de las IA en el contexto de la educación universitaria, así como anticipar el enfoque futuro que representará el uso de estas tecnologías emergentes en los procesos pedagógicos para obtener una comprensión más clara y precisa del contexto actual. (Aguilar et al, 2023).

Los sistemas de inteligencia artificial generativos permiten crear textos o imágenes a través de las demandas de los usuarios. Estas demandas o instrucciones se canalizan a través de los prompts.

Un prompt es una instrucción, pregunta, solicitud o texto que se le da a un programa de inteligencia artificial (IA) para que genere un resultado.

En el contexto de la IA, el prompt es un fragmento de texto que se proporciona a un modelo de lenguaje para influir en la generación de un texto. Por ejemplo, en un chatbot, el prompt es una solicitud que guía la interacción entre el usuario y el sistema de IA. Un prompt bien diseñado debe ser:

- Claro y conciso: Evita ambigüedades y formula tu solicitud de manera directa.
- Específico: Proporciona detalles concretos sobre lo que deseas que se genere.
- Contextualizado: Ofrece suficiente información de fondo para que el modelo comprenda tu petición.
- Bien estructurado: Organiza tus ideas de forma lógica y coherente.
- Orientado a resultados: Define claramente el objetivo que quieres alcanzar.

Aunque los GPTs y su capacidad para generar texto automáticamente han estado disponibles para
los investigadores desde 2018, el lanzamiento de ChatGPT fue novedoso por su acceso gratuito a

través de una interfaz de uso fácil, lo que significa que cualquier persona con acceso a Internet podría
explorar la herramienta. El lanzamiento de ChatGPT causó conmoción en todo el mundo y rápidamente
impulsó a otras empresas tecnológicas transnacionales a readecuarse al nuevo paradigma, sucediendo lo
mismo con numerosas start-ups.

Modelos de IAGen de texto

Los sistemas de IAGen de texto suelen denominarse modelos de lenguaje de gran tamaño, o LLM.

Nombre de la app	Características
Chat GPT	ChatGPT está basado en GPT-3, desarrollado por OpenAI. Esta fue la tercera evolución de su GPT; la primera fue lanzada en 2018 y la más reciente, la GPT-4, en marzo de 2023
Alpaca	Una versión depurada del Llama de Meta, de la Universidad de Stanford, que apunta a abordar la información falsa, los estereotipos sociales y el lenguaje tóxico de los LLM.
Bard	Un LLM de Google, basado en sus sistemas LaMDA y PaLM 2, con acceso a Internet en tiempo real, lo cual le permite ofrecer información actualizada. Ver https://bard.google.com
Chatsonic	Desarrollado por Writesonic, está basado en ChatGPT y rastrea los datos directamente desde Google. En consecuencia, tiene menos posibilidades de generar respuestas incorrectas. Ver https://writesonic.com/chat
Ernie (conocido también como Wenxin Yiyan 文心一言)	Un LLM bilingüe de Baidu, aún en desarrollo, que integra extensos conocimientos con conjuntos de datos masivos para generar texto e imágenes Ver https://yiyan.baidu.com/welcome
Hugging Chat	Creado por Hugging Face, que puso énfasis en la ética y la transparencia durante todo su desarrollo, entrenamiento e implementación. Todos los datos utilizados para entrenar sus modelos son de código abierto. Ver https://huggingface.co/chat
Jasper	Un paquete de herramientas y IPA que, por ejemplo, puede ser entrenado para escribir en el estilo particular preferido de un usuario. También puede generar imágenes.
Llama	Un LLM de código abierto de Meta que requiere menos potencia computacional y menos recursos para probar nuevos enfoques, validar el trabajo de otros y explorar nuevos casos de uso. Ver https://ai.facebook.com/blog/large-language-model-llama-meta-ai
Open Assistant	Un sistema de código abierto diseñado para que cualquier persona con conocimientos suficientes pueda desarrollar su propio LLM. Ha sido creado a partir de datos de entrenamiento recogidos por voluntarios. Ver https://open-assistant.io
Tongyi Qianwen (通义千问)	Un LLM de Alibaba que puede responder a indicaciones en inglés o chino. Está siendo integrado al paquete de herramientas de negocios de Alibaba. Ver https://www.alizila.com/alibaba

	-cloud-debuts-generative-ai-model-for-corporate-users
YouChat	Un LLM que incorpora funciones de búsqueda en tiempo real para brindar contexto y perspectivas adicionales a fin de generar resultados más precisos y confiables. Ver https://you.com

Tabla 1: Modelos de IAGen de texto – Fuente: Guía para el uso de IA generativa en educación e investigación (UNESCO, 2023).pp.7

Si bien en la Tabla 1 se describen los principales IAgen de texto, la velocidad de transformación de estas aplicaciones es exponencial. Por ejemplo, las herramientas basadas en LLM están siendo integradas en otros productos, como los navegadores web. Las extensiones para Chrome basadas en ChatGPT incluyen las siguientes:
WebChatGPT : Permite que ChatGPT acceda a Internet para mantener conversaciones más precisas y actualizadas.
Compose AI: Autocompleta oraciones en correos electrónicos y en otros lugares.
TeamSmart AI:Ofrece un "equipo de asistentes virtuales".
Wiseone: Simplifica la información en línea.

Modelos de IAGen de imágenes
Los modelos de IAGen de imágenes y de música suelen utilizar otro tipo de RNA conocidas como redes
generativas antagónicas (RGAs), que también pueden combinarse con autocodificadores variacionales. Las RGAs constan de dos partes (dos "adversarios"), el "generador" y el "discriminador". En el caso de las RGAs de imágenes, el generador crea una imagen aleatoria en respuesta a un prompt, y el discriminador intenta distinguir entre la imagen generada y las reales.
Posteriormente, el generador utiliza el resultado del discriminador para ajustar sus parámetros y crear otra
imagen. El proceso se repite, quizás miles de veces, y el generador crea imágenes cada vez más realistas que el discriminador distingue progresivamente menos de las reales.



Imagen 1: Imágenes generadas utilizando IAgen. Fuente: Elaboración propia

.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

En este contexto se presentan los objetivos planteados y se detallan los procedimientos implementados para alcanzarlos:

- *Relevar, indagar y seleccionar las apps desarrolladas utilizando IA que tengan*

aplicación en las distintas cátedras de la FIUNLZ.

Se relevarán todas fuentes bibliográficas primarias y secundarias, a fin de reconocer las distintas herramientas de IA.

- *Diseñar objetos de aprendizaje utilizando herramientas de IA.*

Desarrollar e implementar propuestas pedagógicas adecuadas para las carreras de ingeniería. Diseño de rúbricas ajustadas a las propuestas.

- *Aplicar y medir el resultado de aprendizaje de los OA diseñados.*

Una vez desarrollados los OA, se aplican en las distintas cátedras y evalúa el resultado del aprendizaje a través de rúbricas.

3. RESULTADOS OBTENIDOS Y ESPERADOS

Desde la perspectiva del Instituto de Investigaciones de Tecnología y Educación (IITE), el presente proyecto de investigación, propone, desde un marco teórico inscripto en el paradigma de la innovación educativa y la enseñanza y evaluación por competencias, evaluar experiencias desarrolladas utilizando herramientas de IA aplicables a las distintas cátedras, para proponer un modelo que contribuya a la eficiencia de la enseñanza de por competencias en carreras de ingeniería, a través de modelos híbridos (presenciales y virtuales), favoreciendo el aumento de la retención.

4. FORMACIÓN DE RECURSOS HUMANOS

El proyecto prevé la formación de un becario EVC-CIN para el período 2024 - 2025.

5. BIBLIOGRAFÍA

-Aguilar, G. M. F., Gavilanes, D. C. A., Freire, E. M. A., & Quincha, M. L. (2023). Inteligencia artificial y la educación universitaria: Una revisión sistemática. Magazine de las Ciencias: *Revista de Investigación e Innovación*, *8(*1), 109-131.

- Consenso de Beijing sobre la inteligencia artificial y la educación(UNESCO, 2019) En: https://www.unesco.org/es/digital-education/artificial-intelligence

-García Peñalvo, F. J., Llorens Largo, F., & Vidal García, F. J. (2024). La nueva realidad de la educación ante los avances de la inteligencia artificial generativa. RIED. *Revista iberoamericana de educación a distancia.*

- Guidance for generative AI in education and research (2023). Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. En : https://unesdoc.unesco.org/ark:/48223/pf0000389227

-Holmes, W., & Miao, F. (2023). *Guidance for generative AI in education and research*. UNESCO Publishing.

-González, C. S. (2004). Sistemas inteligentes en la educación: una revisión de las líneas de investigación y aplicaciones actuales. RELIEVE. *Revista Electrónica de Investigación y Evaluación Educativa*, *10*(1), 3-22.

-Heredia, M. O. T., Correa, Y. K. D., & Carballo, E. V. (2023). Inteligencia artificial y educación: nuevas relaciones en un mundo interconectado. *Estudios del desarrollo social: Cuba y América Latina*, *11*(2), 312-328.

Diseño de objetos de aprendizaje utilizando metodologías activas, que favorezcan la retención de los estudiantes de la Facultad de Ingeniería de la Universidad Nacional de Lomas de Zamora, en el marco de un plan de estudios por competencias.

Minnaard Claudia, Torres Zulma

Instituto de Investigaciones de Tecnología y Educación (IITE)

Facultad de Ingeniería – Universidad Nacional de Lomas de Zamora

Buenos Aires, Argentina

minnaardclaudia@gmail.com, iite@ingenieria.unlz.edu.ar

RESUMEN

El proyecto se desarrolla en el ámbito del Instituto de Investigaciones en Tecnología y Educación IIT&E de la Facultad de Ingeniería - UNLZ. Entre los objetivos del Instituto se encuentran el interés por contribuir al mejoramiento de la enseñanza y a la formación de competencias de egreso, en particular en carreras científico tecnológicas, mediante la incorporación de tecnologías de la información y la comunicación (TIC) y el diseño de objetos de Aprendizaje. En el presente trabajo se describen los objetivos del proyecto, asi como las acciones realizadas para este fin.

CONTEXTO

La retención de alumnos en las carreras de ingeniería es uno de los temas de mayor preocupación en las universidades. Brito (2011) considera que "la primera aproximación al desarrollo de un modelo de "predicción" para mejorar la retención la hizo Vincent Tinto quien enfatiza variables tales como el respaldo familiar, herramientas y habilidades previas a la universidad, escolaridad previa, compromiso y metas institucionales, integración académica, e integración social".

Son diversos los autores que destacan que el fenómeno del abandono escolar en el nivel universitario ha cobrado gran interés a nivel mundial, son varios los estudios que se dedican a identificar los distintos factores que inciden en dicho fenómeno; sin embargo, la determinación de estrategias de retención, su planificación y puesta en práctica no son tan frecuentes y constituyen aún, todo un desafío. Mastache y Goggi (2017) exponen que una sólida formación envuelve desempeñarse de manera flexible en situaciones profesionales complejas, desarrollando la capacidad de resolución de problemas, a partir de la articulación entre el pensamiento divergente o ingenio y el lógico-matemático. Sostienen que esto sólo es posible a partir de una enseñanza que genere actividades centradas en el aprendizaje, capaces de desarrollar la comprensión y de desplegar procesos que favorezcan la reflexión. Consideran que el desafío, entonces, es pensar una enseñanza que proponga experiencias formativas, que se constituyan en oportunidades para el desarrollo de sujetos autores de su aprendizaje y de su desarrollo profesional. Para ello, se requieren propiciar y sostener estrategias pedagógico-didácticas tales como de la articulación de contenidos de diferentes asignaturas.

Los objetivos del proyecto son los siguientes:

- Establecer relaciones entre desarrollo de competencias en estudiantes de ingeniería, técnicas de enseñanza, estrategias utilizadas para su evaluación y la retención.

- Diseñar objetos de aprendizaje que favorezcan la retención utilizando las metodologías activas.

- Identificar factores clave en el diseño objetos de aprendizaje que aumenten la retención.

- Diseñar COCOs (Clases Online Complementarias) para unidades temáticas de Matemática, Probabilidad y Estadística, Medios de representación I y II.

1. INTRODUCCIÓN

El aprendizaje por competencias está orientado al desarrollo de las habilidades y destrezas necesarias para un adecuado desenvolvimiento en la sociedad. Para desarrollar las habilidades y destrezas necesarias en cada competencia, es muy importante que los docentes tengan acceso a los recursos didácticos adecuados, los cuales puedan adaptarse a las diferentes necesidades educativas de los alumnos. (Morgado, Peñalvo, Ortuño, & Hidalgo. (2015), Comoglio y Minnaard (2020), Minnaard et al (2020), Minnaard et al (2021), Minnaard y Minnaard (2022)).

La formación basada en competencias requiere un enfoque amplio e integrado de todos los aspectos que hacen que al trabajo universitario e implica la participación activa y la interacción de todos los involucrados en el proceso de aprendizaje

En el enfoque de enseñanza por competencias el trabajo curricular se basa en "identificar con claridad las prácticas implícitas o explícitas que se tienen respecto a la formación, con el fin de tomar conciencia de ellas, modificarlas (si es necesario) y buscar generar las condiciones que lleven a tener personas con alto compromiso ético, autorrealización, emprendimiento e idoneidad para afrontar los diferentes retos del contexto" (Tobón Tobón, S., 2011).

En este sentido, Caraballo Carmona, C., Meléndez Ruiz, R., & Iglesias Triana, L. (2019), consideran que "una nueva forma para el logro de estos objetivos tiene sus bases en las competencias, lo cual implica, saber planear por competencias, enseñar por competencias y evaluar por competencias. Competencia, no significa habilidad para ejecutar una acción, esta no se limita a los procedimientos, sino que además va acompañada necesariamente de elementos teóricos y actitudinales.

Por su parte, Amaya Chávez (2021) considera que para llevar a cabo la formación y desarrollo de competencias en los estudiantes es necesaria la aplicación de las metodologías activas. Entre estas se destacan: Aprendizaje colaborativo, Estudio de casos, Aprendizaje basado en proyectos, Aprendizaje basado en problemas. Mas recientemente, se tienen el Aula invertida o Flipped classroom, Gamificación, Visual thinking, Design thinking y Studio based learning (Palma et al, 2017).

Las metodologías activas se basan en que el alumno a través de la guía del docente asuma con responsabilidad y autonomía su proceso de aprendizaje, y que se refleje en la mejora de sus desempeños tanto conceptuales (conocimientos, teorías y leyes), procedimentales (habilidades y destrezas), así como actitudinales (intereses, motivos y modos de actuación), por lo que es fundamental generar un replanteamiento metodológico que permita el desarrollo de actividades en donde el centro sea la construcción de sentidos y de significados a partir de un proceso activo y constructivo por parte del estudiante, tanto de manera individual como de forma colaborativa

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El aprendizaje eficaz requiere que los alumnos operen activamente en la manipulación de la información a ser aprendida, pensando y actuando sobre ello

para revisar, expandir y asimilarlo. Por lo tanto y desde una perspectiva didáctica, estamos construyendo competencias.

Para alcanzar los resultados esperados en el presente proyecto de investigación, y en función de los objetivos planteados, se requiere trabajar con datos provenientes de distintas fuentes, los que serán sometidos a distintos tipos de análisis.

A continuación, se presentan los objetivos planteados y se desarrolla la explicación solicitada

- *Establecer relaciones entre desarrollo de competencias en estudiantes de ingeniería, técnicas de enseñanza, estrategias utilizadas para su evaluación y la retención.*
  Encuestas, entrevistas y focus group a los alumnos ingresantes. Interpretación y análisis de discurso social de los resultados de las entrevistas y focus group. Análisis estadístico (descriptivo y correlacional). Relevamiento de estrategias innovadoras que los docentes integran a sus prácticas y sistematización de casos describiendo: objetivos, metodología desarrollada, resultados esperados y alcanzados.

- *Diseñar objetos de aprendizaje que favorezcan la retención utilizando las metodologías activas*.
  Desarrollar e implementar propuestas pedagógicas adecuadas. Diseño de rúbricas ajustadas a las propuestas.

- *Identificar factores clave en el diseño objetos de aprendizaje que aumenten la retención.*
  Relevamiento de estrategias innovadoras que los docentes integran a sus prácticas y sistematización de casos describiendo: objetivos, metodología desarrollada, resultados esperados y alcanzados.

- *Diseñar COCOs (Clases Online Complementarias) para unidades temáticas de Matemática, Probabilidad y Estadística, Medios de representación I y II.*
  Desarrollar e implementar propuestas pedagógicas adecuadas. Diseño de rúbricas ajustadas a las propuestas

3. RESULTADOS OBTENIDOS Y ESPERADOS

Los destinatarios directos de los resultados del presente programa son estudiantes y docentes de las carreras de Ingeniería Mecánica, Industrial, Ferroviaria y Mecatrónica de la FI UNLZ , a través de las aplicaciones y conocimientos que se generen en el ámbito de las cátedras y espacios institucionales en los que se desempeñan los miembros del grupo de investigación.

4. FORMACIÓN DE RECURSOS HUMANOS

La formación de Recursos Humanos incluye la formación de 2 becarios del programa EVC-CIN para el período 2024-2025

5. BIBLIOGRAFÍA

Amaya Chávez, D. (2021). Diseño, análisis, y evaluación de un método de aprendizaje basado en problemas (abp) mediado por tecnologías interactivas desarrollado con estudiantes de ingeniería.

Caraballo Carmona, C., Meléndez Ruiz, R., & Iglesias Triana, L. (2019). Reflexiones acerca del concepto competencias y aprendizaje por competencias en las instituciones de educación superior y su incidencia en el aprendizaje de las matemáticas. Opuntia Brava, 11(1), 297-307.

Comoglio, M. & Minnaard, C. (2020) Desafíos y adaptaciones en el proceso de enseñanza en las distintas cátedras de la Facultad de Ingeniería de la Universidad Nacional de Lomas de Zamora en tiempos de pandemia. V Encuentro Centros CIC, Buenos Aires, Argentina.

Mastache y Goggi (2017) El Comité de Mejora una experiencia colaborativa en Ingeniería. Revista Argentina de Educación Superior N° 14.

Minnaard, C. & Minnaard, V. (2014). Evaluación por competencias en Entornos Virtuales de Aprendizaje. XVI Workshop de Investigadores en Ciencias de la Computación. En: http://sedici.unlp.edu.ar/bitstream/handle/10915/43712/Documento_completo.pdf?sequence=1&isAllowed=y

Minnaard,A.; Iravedra, C.; Garrido, G.& Gonzalez, B.(2020) Enseñanza por competencias: la experiencia de los COCOs. VII Jornadas Nacionales y III Latinoamericanas de Ingreso y Permanencia en Carreras Científico – Tecnológicas

Minnaard, C.;Comoglio, M.; Iravedra, C.;Garrido, G. & Gonzalez, B.(2020) Enseñanza por competencias: la experiencia de los COCOs. 18th LACCEI International Multi-Conference for Engineering, Education, and Technology: "Engineering, Integration, and Alliances for a Sustainable Development" "Hemispheric Cooperation for Competitiveness and Prosperity on a Knowledge-Based Economy", 29-31 July 2020, Buenos Aires, Argentina.

Minnaard,C.; Minnaard, V. & Torres, Z. (2021) Adaptaciones en el proceso de enseñanza en la cátedra de Probabilidad y Estadística en tiempos de pandemia. XVI Congreso Internacional de Ciencias Sociales Interdisciplinares, Oxford, Gran Bretaña.

Minnaard, V. & Minnaard, C. (2022) Design and Validation of an Instrument to Analyze Mobile Applications that Favor mLearning: Mobile Applications Subject to Analysis. XVIII International Conference on Technology , Knwoledge and Society. Universidad Changhua, Taiwan.

Morgado, E. M., Peñalvo, F. G., Ortuño, R. C., & Hidalgo, C. A. (2015). Desarrollo de competencias a través de objetos de aprendizaje. Revista de Educación a Distancia, (36).

Palma Luengo, M. et al (2017) Metodologías activas para la formación universitaria. Universidad de Bio Bio.

Tobón, S. (2011). Formación basada en competencias. Pensamiento complejo, diseño curricular y didáctica. 2 ed. Bogotá: ECOE Ediciones

ENSEÑAR APLICANDO INNOVACIONES TECNOLÓGICAS DE INTELIGENCIA ARTIFICIAL GENERATIVA

Mg. Mariela Delauro (mdelauro), Lic. Mariela Sánchez (msanchez), Esp. Mabel Compagnoni (mcompagnoni) @unlam.edu.ar

Departamento de Ingeniería e Investigaciones Tecnológicas

Universidad Nacional de La Matanza

Área de Interés: Tecnología Informática Aplicada en la Educación

RESUMEN

El presente trabajo analiza la capacitación implementada en la Universidad Nacional de La Matanza (UNLaM), específicamente desde el Departamento de Ingeniería e Investigaciones tecnológicas (DIIT) mediante la cual, los docentes abordan la aplicación de las herramientas tecnológicas provistas por Inteligencia Artificial (en adelante IA) dentro del ámbito educativo.
Actualmente la irrupción de la IA Generativa ha puesto en evidencia la necesidad de que los docentes se preparen para aprovechar las potencialidades de las herramientas innovadoras brindadas por ella en el proceso de enseñanza. Para esto, se diseñaron dos cursos de capacitación tendientes al saber – hacer, donde se facilita el uso de la tecnología (IA) aplicada en la educación, mediante tutoriales y ejemplos concretos, que permitan valorar la posibilidades y ventajas de aplicación en la práctica pedagógica cotidiana.
Se implementó una encuesta PINA (Positivo, Interesante, Necesario, Aplicabilidad) para ser completada por los cursantes de la capacitación, la que mostró un alto nivel de satisfacción de los docentes respecto a la aplicación de la IA en educación.

Palabras Claves: *Inteligencia artificial; Enseñanza; Innovación; Capacitación; Docencia*

CONTEXTO

La educación se caracteriza por la transformación en la construcción del conocimiento y en las formas de transferencia de estos nuevos saberes, por eso las universidades están reconfigurando sus escenarios de enseñanza ante retos como la incorporación de la IA en el ámbito educativo.

La inteligencia artificial (IA) está transformando la manera en que vivimos, trabajamos y aprendemos. Para los docentes, comprender y saber integrar estas tecnologías en el aula es crucial para preparar a los estudiantes para el futuro.

Actualmente, hay muchas aplicaciones desarrolladas con IA destinadas al ámbito educativo. Se considera necesario que los docentes se preparen para usar estas aplicaciones en el quehacer cotidiano del ejercicio de su labor pedagógica e implementarlas dentro del modelo TPACK, donde la Tecnología del Aprendizaje y la Comunicación se integra a los procesos de enseñanza y aprendizaje para favorecer la construcción del conocimiento y facilitar el acompañamiento del trayecto formativo de los estudiantes. Además, los participantes pueden

identificar las ventajas de la IA en Educación. Entre otras:

- *Automatización de Tareas*: reduciendo carga laboral administrativa y de evaluación.
- *Creación de contenido inteligente*: lecciones digitales y recursos educativos variados.
- *Asistencia permanente:* chatbots y apoyo educativo.

El propósito de este trabajo es analizar las experiencias académicas tras la implementación de estos cursos.

1. INTRODUCCION

La educación superior ha experimentado una notable evolución en las últimas décadas, impulsada en gran medida por los avances tecnológicos. Entre estos avances, la Inteligencia Artificial destaca como una herramienta revolucionaria que promete transformar la educación tradicional en un modelo más personalizado y adaptado a las necesidades individuales de los estudiantes.

Para algunos autores como Barberà, E, Badia, A. (2004:12) "Los conocimientos tecnológicos no aparecen como un objetivo prioritario porque son un medio para el propio proceso de enseñanza y aprendizaje, pero al serlo, paradójicamente, se convierten en un foco principal de atención por parte del profesor, puesto que representan el canal de comunicación y la entrada a la relación didáctica".

En el ámbito educativo la IA permite al profesor diseñar y planificar tareas, recursos, actividades y evaluaciones de aprendizaje, optimizando su labor para adaptarla a las necesidades de los estudiantes.

Para los docentes, la IA ofrece la oportunidad de automatizar algunas tareas, como diseñar una autoevaluación, la corrección de un ensayo o el desarrollo de una actividad práctica, proporcionando una retroalimentación detallada y específica.

Además, los chatbots/asistentes educativos son otra herramienta que permite liberar el tiempo de los docentes, si están diseñados a partir de una temática o campo de acción, proporcionan respuestas y guían al estudiante en la lectura comprensiva de los contenidos.

Los tutores virtuales, basados en IA, ofrecen apoyo adicional a los estudiantes, como explicaciones, resolución de preguntas y ejercicios de práctica, algunos pueden incluir recomendación de contenido que se adapten a los intereses y habilidades de cada estudiante.

La elaboración de materiales y recursos educativos, proporcionan alternativas de enseñanza. Consideramos que la relación docente-estudiante es bidireccional, es decir, hay intención de quien enseña y de quien aprende.

Por ello, diseñar un entorno de aprendizaje interactivo que capte la atención de estudiantes y los motive a participar y seguir aprendiendo mejora la comprensión de los conceptos y la retención de lo estudiado.

Por supuesto, la implementación efectiva de la IA requiere de un proceso de planificación y capacitación, en busca del beneficio tanto para docentes como estudiantes. Siempre teniendo presente que la IA no reemplaza la experiencia, conocimientos y habilidades humanas, sólo es un complemento de utilidad para ayudar en la tarea y mejorar el proceso de enseñanza y aprendizaje.

González, Milillo y Tarasow (2024), proponen una sistematización - escalera - del uso de la IA para incorporarla al proceso de enseñanza, así pues, la conciben como una herramienta, capaz de generar un producto o una respuesta simple. Luego incorporan el chatbot en la toma de decisiones, lo que permite potenciar la creatividad del docente para el diseño didáctico.

De acuerdo con Mireia Ribera y Díaz Montesdeoca (2024), entendemos que un

Chatbot (o Bot conversacional) es un programa informático diseñado para simular una conversación en lenguaje natural con seres humanos a través de un canal de texto, voz o incluso imágenes, además tiene la capacidad de procesar y comprender las preguntas y solicitudes del usuario proporcionando respuestas relevantes y coherentes.

Por ello la IA puede analizar/aprender de usuarios docentes para recomendar material relevante. Si su uso es a modo de "asistente", la IA conversacional puede dar apoyo en la etapa de planificación y organización, respondiendo a preguntas que adicionan explicaciones o ejemplos prácticos. Por último, como soporte a la tarea docente una IA generativa puede incluir nuevas ideas sobre actividades a realizar dentro del aula, en relación con una temática específica. Rey Valzacchi (2024) coincide con esto y enfatiza que "...tienen la capacidad de personalizar el proceso de aprendizaje, evaluar el progreso de los estudiantes de manera instantánea y facilitar el acceso una gran cantidad de recursos educativos".
Jorge Rey Valzacchi (2024), realiza una certera distinción entre los tipos de prompts: informativos, secuenciales, condicionales, reflexión, simulación, creativos.

2. LINEA DE INNOVACIÓN

Se implementaron dos cursos taller para la integración pedagógica, con estos objetivos:

- Reflexionar sobre el uso pedagógico de los Chatbots - Asistentes, sus posibilidades y limitaciones.
- Identificar las ventajas del uso de herramientas digitales de Inteligencia Artificial para la creación de recursos y generación de actividades.
- Determinar el valor de las herramientas de generación de recursos didácticos como un medio facilitador dentro de la tarea docente.
- Reflexionar críticamente sobre el uso de las distintas aplicaciones de la Inteligencia Artificial desde el punto de vista ético.

La propuesta concreta se estructura de la siguiente manera:

- ***Curso 1. Uso de Chatbots - Asistentes***: sus ventajas, posibilidades, desafíos y aspectos para tener en cuenta para incluirlo dentro de los procesos de enseñanza y aprendizaje. Se plantearán alternativas para su aplicación, dejando debates para la reflexión.

- ***Curso 2. Creación de presentaciones y recursos para la enseñanza:*** reducción de la carga laboral de los docentes si se usan adecuadamente. Propuestas de distintas herramientas de aplicación concretas en aulas (presenciales, virtuales o híbridas).

- Además, se implementó la creación de un "***Banco de Recursos***" en línea para compartir con todos los cursantes el cual incluye los nombres de las nuevas herramientas IA de innovación tecnológica para aplicar en el ámbito educativo.

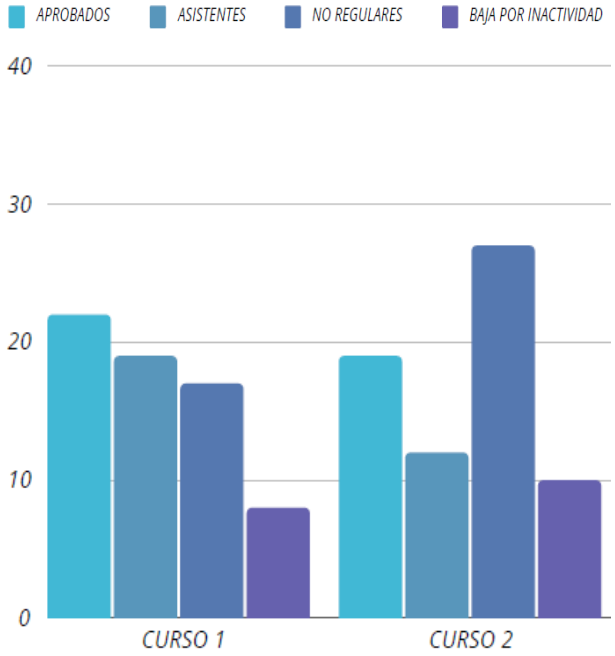
3. RESULTADOS OBTENIDOS

En la encuesta PINA (Positivo, Interesante, Necesario, Aplicabilidad) que respondieron los cursantes se destacan los siguientes aspectos:

- *Facilidad de uso:* Los participantes encuentran las herramientas de IA intuitivas, lo que facilita su aprendizaje y aplicación en sus clases.
- *Mejora en la creación de contenido*: Utilizan la IA para mejorar los prompts y obtener resultados más específicos. Las herramientas ayudan en la creación de presentaciones y videos que refuerzan el aprendizaje del estudiante. También posibilitan la gamificación en el aula como estrategia de aprendizaje.
- *Curiosidad por tutoriales y experimentación:* Reconocen la utilidad

- de los tutoriales para aprender a usar las aplicaciones. La práctica mediante prueba y error es fundamental para dominar las herramientas.
- *Eficiencia y organización*: Las herramientas permiten simplificar el trabajo y generar interés entre los estudiantes. Se destaca el ahorro de tiempo en la tarea docente y acceso a recursos visuales variados y creativos.
- *Uso de herramientas específicas:* Aplicaciones como Gemini, GPT, Fliki y Gamma, aunque se expresa un interés por probar otras disponibles. Enfatizan la importancia de analizar críticamente la información obtenida.
- *Interacción con los estudiantes:* Hay una expectativa por compartir lo aprendido con sus estudiantes, aumentando su interés y participación.

Se obtuvieron los siguientes resultados cuantitativos relativos a los docentes cursantes:



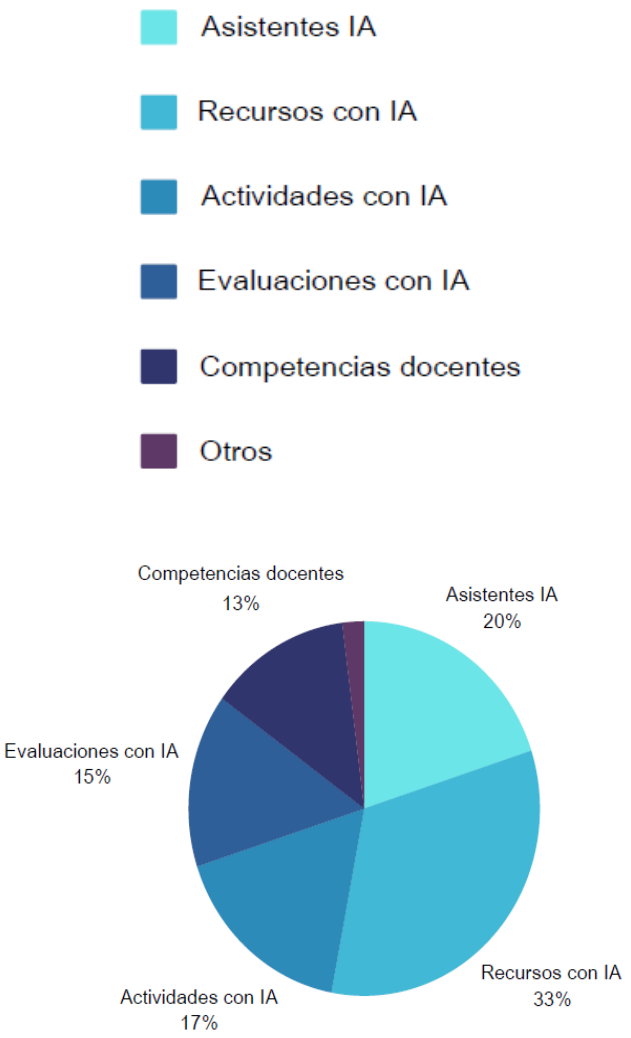
4. FORMACION DE RECURSOS HUMANOS

La implementación de la propuesta generó una importante convocatoria y tuvo una gran aceptación por parte de los docentes, 75 inscripciones a los cursos propuestos distribuidos de la siguiente manera:



Cabe destacar, que la mayoría de los docentes inscriptos pertenecen al DIIT, pero también docentes de otros departamentos participaron de la capacitación. Además, lo importante es que el 77% se inscribió en todos los cursos.

A continuación, se detallan los resultados de las respuestas de los 42 docentes del DIIT que respondieron a la consulta formulada respecto de cuáles capacitaciones serían de su interés para realizar a futuro:



5. BIBLIOGRAFIA

• Barberà, E, Badia, A. (2004) Educar con aulas virtuales. Madrid: A. Machado Libros. Cap. I y II

• González, N.; Milillo, C.; Tarasow, F.; (2024). "La escalera de la inteligencia artificial en educación: orientaciones para la incorporación de IA en la tarea docente." Recuperado de: https://pent.flacso.org.ar/producciones/escalera-inteligencia-artificial-educacion

- Rey Valzacchi, Jorge (2024) Aprendizaje ilimitado: Potenciando la educación con Chat GPT y DALL-E / Una exploración pragmática de la IA en la educación. URL descarga gratuita: https://acortar.link/FaX7Tw

- Rey Valzacchi, Jorge (2024) "Chatbots educativos – Inteligencia artificial transformando el aprendizaje – El nuevo paradigma de la enseñanza personalizada". URL descarga gratuita: https://acortar.link/EBi35J
- Mireia Ribera, O.; Díaz Montesdeoca (coord.) (2024). "ChatGPT y educación universitaria: posibilidades y límites de ChatGPT como herramienta docente." Ed. Octaedro. Barcelona.

Estrategias para la Mejora Continua del Sistema de Aseguramiento de la Calidad de la UNdeC

Fernanda CARMONA[1], Emanuel Agustín IRRAZABAL[3], Mara Lía Inés ROVERO[1], Ana Carolina MANRIQUE [1], Paula María RIVAS [2], Nelson Javier CORDOBA VILTE[1], Amadeo Jesús CLARAMONTE [1], Lucas DI BENEDETTO[2]

[1]Departamento de Ciencias Básicas y Tecnológicas,
[2]Departamento de Ciencias Sociales, Jurídicas y Económicas, Universidad Nacional de Chilecito 9 de Julio 22, Chilecito, La Rioja, Argentina; [3] Grupo de Investigación en Calidad de Software, Facultad de Ciencias Exactas y Naturales y Agrimensura, Universidad Nacional del Nordeste, 25 de Mayo 868, Corrientes, Argentina

{fbcarmona, mrovero, amanrique, privas }@undec.edu.ar, {emanuelirrazabal, javiercordoba0899, claramontejesus, lucasdibenedetto7}@gmail.com

RESUMEN

Gestionar la Calidad implica un conjunto de acciones y herramientas coordinadas para dirigir y controlar una organización de acuerdo a ciertos requisitos y estándares. Respecto a la calidad en la educación superior existen varias acepciones, suele definirse como el grado en el que un conjunto de rasgos diferenciadores inherentes a la educación superior cumple con una necesidad o expectativa establecida, también como la propiedad de una institución o programa que cumple los estándares previamente establecidos por una agencia u organismo de acreditación. La norma ISO 21001:2018 alineada con la norma ISO 9001 que promueve la adopción de un enfoque a procesos al desarrollar, implementar y mejorar la eficacia de un Sistema de Gestión de Organizaciones Educativas, para aumentar la satisfacción del estudiante y de otros beneficiarios mediante el cumplimiento de los requisitos de estos. El crecimiento y desarrollo de la UNdeC, requiere de un proceso de identificación de las actividades desarrolladas y de los procesos relacionados con el objeto de implementar acciones preventivas, correctivas y de análisis para la satisfacción de los destinatarios finales de los servicios que la institución brinda. Esta línea de investigación tiene por objeto el análisis y estudio de los requisitos definidos por la norma ISO 21001:2018 para la definición y desarrollo de estrategias que permitan implementar la mejora continua del Sistema de Aseguramiento de la Calidad de la UNdeC.

Palabras clave: Gestión de Calidad, Enfoque a procesos, Educación Superior, Normas ISO.

CONTEXTO

La Ley N° 24.521 de Educación Superior establece que la autoevaluación institucional y la evaluación externa apuntan a la mejora y el aumento en la calidad del funcionamiento de las universidades y del cumplimiento de sus propósitos. Dispone que las instituciones universitarias deben asegurar el funcionamiento de instancias internas de evaluación institucional, que tendrán por objeto analizar los logros y dificultades en el cumplimiento de sus funciones, así como sugerir medidas para su mejoramiento; instaura la tutela del bien público en actividades que puedan ser afectadas por intervenciones profesionales, estableciendo para ello mecanismos de acreditación periódicas; entre otras funciones propias de las instituciones que conforman la Educación Superior, establece la de garantizar crecientes niveles de calidad y

excelencia en todas las opciones institucionales del sistema.

En este contexto desde el año 2005 la UNdeC viene transitando diversos procesos de acreditación de carreras y en el 2021 firma el acuerdo con CONEAU para implementar el proceso de evaluación institucional actualmente finalizado.

Con el objeto de capitalizar la experiencia adquirida en estos procesos y de sistematizar los procedimientos y aprendizajes de futuros procesos, en abril del año 2018 la UNdeC crea la Dirección de Evaluación y Acreditación, RR Nº 203/18, con la responsabilidad primaria de asegurar el adecuado desarrollo de las instancias de evaluación institucional y de acreditación de carreras, con el fin de garantizar la calidad de la oferta académica y de la gestión institucional y su mejora continua.

La Resolución 2597/2023 del ME establece que las instituciones universitarias deberán garantizar, en cada una de ellas y en el marco de sus pautas organizativas y de gobierno, el funcionamiento de sistemas institucionales de aseguramiento de la calidad (SIAC) a cargo de las unidades de gestión correspondientes. Los SIAC serán, sin perjuicio de otras funciones que la institución les asigne, responsables de promover y/o fortalecer la cultura de evaluación de calidad; administrar y desarrollar las estrategias, instrumentos y procesos de evaluación institucional; realizar la certificación de carreras; promover planes de mejora acordes con las evaluaciones realizadas y en el marco del respectivo plan de desarrollo institucional.

El logro de estos requerimientos implica perfeccionar mecanismos y dispositivos de autorregulación de las instituciones universitarias en sus procesos de mejora de la calidad. Es así que resulta recomendable la institucionalización de procesos, dispositivos y mecanismos estables, orgánicos, regulares y sustentables para la evaluación, planeamiento y coordinación de los procesos de mejora en el cumplimiento de las funciones institucionales, conforme a especificaciones y estándares de calidad nacional e internacional.

En Argentina, la Comisión Nacional de Evaluación y Acreditación Universitaria (CONEAU) es el organismo descentralizado que funciona con la finalidad de contribuir al mejoramiento de la educación universitaria.

La norma ISO 21001:2018 es la primera norma internacional orientada a la especificación de los requisitos de un Sistemas de Gestión de Organizaciones Educativas (SGOE).

La UNdeC como Institución de Educación Superior se encuentra con el desafío de responder a las nuevas exigencias y competencias relacionadas con la sociedad del conocimiento y con ello a las nuevas actividades pedagógicas y a los requerimientos de acciones más eficientes en la gestión institucional.

El mejoramiento continuo requiere de estrategias, herramientas, planificación de acciones y de despliegue de recursos para el logro eficiente de sus objetivos institucionales. El aseguramiento de la calidad, es la acción de garantizar la calidad de los procesos y resultados en función de criterios y estándares preestablecidos, siendo la base para asegurar la estabilidad del proceso educativo. [7]

La Resolución Rectoral UNdeC N° 777/24 aprueba el financiamiento y ejecución del proyecto “Especificación de procesos bajo la norma ISO 21001:2018 para el fortalecimiento del sistema de aseguramiento de la calidad de la UNdeC.” presentado en la convocatoria UNdeC - Financiamiento para estímulo y desarrollo de la Investigación Científica y Tecnológica, período de ejecución 2024-2025. Este proyecto tiene por objeto el análisis y estudio de los requisitos definidos por la norma ISO 21001:2018 para especificar los procesos

del sistema de aseguramiento de la calidad de la UNdeC con el fin de desarrollar, implementar y mejorar su eficacia y eficiencia.

INTRODUCCIÓN

La ISO 9000 define la calidad como todas las características de un producto o servicio que satisfacen los requisitos del cliente y cumplen con todos los reglamentos y políticas aplicables. [5]

Las organizaciones buscan gestionar la calidad por medio de la implementación de estándares de calidad que aseguren su rentabilidad y la satisfacción del cliente involucrando los esfuerzos de todo el personal y utilizando metodologías que permiten conocer las necesidades particulares del cliente y determinar su percepción de la calidad para traducirlas en especificaciones de las características que debe cumplir el producto o el servicio que se ofrece. En este sentido, la familia de estándares de calidad más utilizada en las administraciones públicas y privadas es la de las normas ISO 9000 que incluye las siguientes: ISO 9000:2015, Sistemas de gestión de la calidad, Fundamentos y vocabulario; ISO 9001:2015, Sistemas de gestión de la calidad, Requisitos; ISO 9002:2016, Sistemas de gestión de la calidad, Directrices para la aplicación de la Norma ISO 9001:2015; ISO 9004:2018, Gestión de la calidad, Calidad de una organización, Orientación para lograr el éxito sostenido.

Gestionar la Calidad implica la definición de un conjunto de acciones y herramientas coordinadas para dirigir y controlar una organización con respecto a la calidad. Según la Norma ISO 9000: 2015 "la Gestión de Calidad puede incluir el establecimiento de las políticas y objetivos de calidad junto con los procesos para lograr estos objetivos de la calidad a través de la planificación de la calidad, el aseguramiento de la calidad, el control de la calidad y la mejora de la calidad"[4]. La Gestión de Calidad incluye:

- Planificación de la calidad: orientada a establecer los objetivos de la calidad y la especificación de los procesos operativos y los recursos relacionados necesarios para lograr los objetivos de la calidad.
- Aseguramiento de la calidad: orientada a proporcionar confianza en que se cumplirán los requisitos de calidad.
- Control de calidad: orientada al cumplimiento de los requisitos de la calidad.
- Mejora de la calidad: orientada a aumentar la capacidad de cumplir con los requisitos de la calidad.

Por su parte, un Sistema de Gestión de Calidad (SGC) comprende actividades mediante las cuales la organización identifica sus políticas, objetivos y determina los procesos y recursos requeridos para lograr sus objetivos. A su vez, gestiona los procesos posibilitando a la alta dirección optimizar el uso de los recursos considerando sus decisiones a largo y corto plazo y abordar las consecuencias previstas y no previstas en la provisión de productos y servicios.[4]

Respecto de la calidad en la educación superior existen varias acepciones, suele definirse como el grado en el que un conjunto de rasgos diferenciadores inherentes a la educación superior cumple con una necesidad o expectativa establecida, se define además como la propiedad de una institución o programa que cumple los estándares previamente establecidos por una agencia u organismo de acreditación. [10]

La evaluación de la calidad en la educación universitaria constituye uno de los temas claves en la agenda de reformas que se comienzan a implementar en este sector hacia mediados de la década del '80 en casi todas las regiones del mundo. En este contexto, en casi todos los países latinoamericanos han surgido agencias públicas o privadas de evaluación cuyos objetivos generales son asegurar y mejorar la calidad de la formación universitaria. En el caso de Argentina, la CONEAU cumple la misión institucional de asegurar y mejorar la calidad de

las carreras e instituciones universitarias que operan en el sistema universitario argentino por medio de actividades de evaluación y acreditación de la calidad de la educación universitaria.

Para lograr una mejora continua, las instituciones de educación deben reconocer a los estudiantes, como la razón principal de su existencia y aplicar políticas que mantengan relaciones positivas con estos, el personal académico, administrativo y todos los miembros que conforman la universidad. [11]

La norma ISO 21001:2018 es una norma de sistema de gestión independiente, alineada con la Norma ISO 9001 que promueve la adopción de un enfoque a procesos al desarrollar, implementar y mejorar la eficacia de un Sistema de Gestión de Organizaciones Educativas (SGOE), para aumentar la satisfacción del estudiante y de otros beneficiarios mediante el cumplimiento de los requisitos de estos.

La comprensión y gestión de los procesos interrelacionados como un sistema contribuye a la eficacia y eficiencia de la organización en el logro de sus resultados previstos. El enfoque a procesos implica la definición y gestión sistemática de los procesos y sus interacciones, con el fin de alcanzar los resultados previstos de acuerdo con la política, los objetivos y el plan estratégico de la organización. La aplicación del enfoque a procesos en un SGOE permite la comprensión y coherencia en el cumplimiento de objetivos; la consideración de los procesos en términos de valor agregado; el logro del desempeño eficaz del proceso; y la mejora de procesos con base en la evaluación de los datos y la información. Una organización necesita planificar e implementar acciones para abordar los riesgos y las oportunidades, esto establece una base para aumentar la eficacia del sistema de gestión de calidad, alcanzar mejores resultados y prevenir los efectos negativos. [6]

LÍNEAS DE INVESTIGACIÓN y DESARROLLO

- Gobernanza y políticas institucionales.
- Calidad académica y administrativa.
- Diseño curricular y planificación académica.
- Indicadores de desempeño alineados con normas de calidad.
- Gestión de Procesos y Evaluación de Desempeño.
- Automatización y digitalización de procesos para el aseguramiento de la calidad.
- Interoperabilidad.
- Preservación digital.
- Acceso y recuperación de la información.

OBJETIVOS

El desarrollo de esta línea de investigación está permitiendo cumplir con los siguientes objetivos:

- Analizar en detalle los requisitos de la Norma ISO 21001:2018 y su aplicación en Instituciones Educativas.
- Realizar un diagnóstico de la situación actual de los procesos de evaluación y acreditación implementados para determinar el grado de cumplimiento de los requisitos establecidos por la Norma ISO 21001:2018.
- Establecer un plan de trabajo para el diseño, especificación y actualización de los procesos del Sistema de Aseguramiento de la Calidad que permita dar cumplimiento a los requisitos de la norma ISO 21001:2019.
- Diseñar el mapa de procesos y especificar los procesos identificados en la norma ISO 21001:2018.
- Definir los indicadores que permitan evaluar la mejora continua en los procesos especificados.

- Analizar la factibilidad de automatización de los procesos y la sistematización de la información generada.

RESULTADOS OBTENIDOS/ESPERADOS

- Identificación de los requisitos establecidos en la norma para la especificación de procesos.
- Diagnóstico de los procesos de evaluación y acreditación implementados para determinar el grado de cumplimiento de los requisitos establecidos por la norma.
- Especificación y actualización de los procesos del sistema de aseguramiento de la calidad.
- Especificación, actualización y formalización de procedimientos, instructivos y registros vinculados a cada proceso.
- Estudio de factibilidad para la automatización de procesos y la sistematización de la información requerida.

FORMACIÓN DE RECURSOS HUMANOS

El equipo está conformado por tres docentes investigadoras del Departamento de Básicas y Tecnológicas, una docente investigadora del Departamento de Ciencias Sociales, Jurídicas y Económicas de la UNdeC, un docente investigador de la Facultad de Ciencias Exactas y Naturales y Agrimensura de la Universidad Nacional de Nordeste. Tres estudiantes avanzados de la UNdeC.

De los integrantes del equipo: una docente investigadora en el área de Ingeniería de Software y Tecnologías en Educación, categoría IV del programa de incentivos, ha ocupado cargos de gestión en diferentes áreas de la UNdeC relacionadas con la gestión académica e institucional; un docente investigador de la UNNE, categoría III del programa de incentivos, Director del Grupo de Investigación en Calidad del Software y de proyectos de investigación relacionados con calidad del software, acreditado como Auditor Jefe Norma ISO 15504 por AENOR y Auditor de personal I+D por la Agencia Española de Normalización. Dos de las docentes investigadoras integrantes del equipo están vinculadas a la Dirección de Evaluación y Acreditación de la UNdeC como soporte a los procesos de evaluación institucional y acreditación de carreras, una de ellas es docente investigadora en el área de Auditoría y Calidad de Software, integrante de comisiones de autoevaluación en los procesos de acreditación de las carreras de la UNdeC. Los estudiantes que forman parte del equipo, son estudiantes avanzados, dos de ellos de la carrera Ingeniería en Sistemas y uno de la carrera Licenciatura en Economía de la UNdeC, los tres desarrollando sus trabajos finales sobre temáticas vinculadas a esta línea de investigación.

La formación y capacidades de cada uno de los miembros del equipo de trabajo y en especial la colaboración de experiencias de docentes investigadores de dos instituciones universitarias fomenta la reflexión, el debate y contraste de diferentes puntos de vistas, fortaleciendo el conocimiento, capacidades, habilidades y destrezas en la utilización de herramientas, métodos y técnicas que forman parte del proceso de investigación.

BIBLIOGRAFÍA

[1]. ARAICA ZEPEDA, Ricardo. Diseño de un Sistema de Gestión y Aseguramiento de la Calidad de las Instituciones de Educación Superior de Nicaragua: Caso de una Universidad pública. 2023. Tesis Doctoral. Universidad Nacional Autónoma de Nicaragua.

[2]. CONEAU. Guía para la implementación y consolidación de sistemas internos de aseguramiento de la calidad universitaria. 2023

[3]. CORDOBA MORENO, Romina Alejandra. Sistema de gestión de la

calidad en la Universidad Nacional de Rosario. Programa de implementación y certificación de un sistema de gestión de la calidad según los requerimientos de la Norma IRAM-ISO 9001. 2023. Tesis de Maestría..

[4]. DE NORMALIZACIÓN, Asociación Española. UNE-ISO 9000 Sistema de gestión de calidad. Fundamentos y vocabulario. 2015.

[5]. GOETSCH, David L.; DAVIS, Stanley B. Quality management for organizational excellence: Introduction to total quality. pearson, 2016.

[6]. ISO 21001:2018. Norma Internacional ISO 21001:2018 traducción oficial ISO, primera edición. 2018.

[7]. PÉREZ, Ivonne Rodríguez. Sistema de gestión de la calidad y su influencia en la calidad educativa. Revista Iberoamericana de Producción Académica y Gestión Educativa, 2015, vol. 2, no 3.

[8]. SALGADO CHILON, Kelly Gabriela. Sistemas de gestión de calidad bajo la norma ISO 21001: 2018, para procesos académicos de la Escuela Académico Profesional de Ingeniería de Sistemas de la Universidad Nacional de Cajamarca. 2024.

[9]. SOLORZANO, Diana; CALIUSCO, María Laura; ARAUJO, Pedro. Identificación de stakeholders para la autoevaluación de calidad en una universidad pública: caso de estudio UTN-FRT. En XV Simposio de Informática en el Estado (SIE 2021)-JAIIO 50 (Modalidad virtual). 2021.

[10]. STRAH, Martín. Los sistemas nacionales de aseguramiento de la calidad de la Educación Superior en Iberoamérica. Comisión Nacional de Evaluación y Acreditación Universitaria (CONEAU) y Organización de Estados Iberoamericanos (OEI), 2020.

[11]. SUNDER M, Vijaya. Constructos de calidad en los servicios de educación superior. Revista internacional de gestión de la productividad y el rendimiento, 2016, vol. 65, núm. 8, pág. 1091-1111.

ESTUDIO DE TÉCNICAS DE INTELIGENCIA ARTIFICIAL PARA LA EVALUACIÓN EN FORMA CONJUNTA DE LA USABILIDAD Y ACCESIBILIDAD DE LAS APLICACIONES

Javier Díaz, Alejandra Schiavoni, Paola Amadeo, Ivana Harari,
Alejandra Osorio
LINTI - Laboratorio de Investigación en Nuevas Tecnologías Informáticas.
Facultad de Informática. Universidad Nacional de La Plata
Calle 50 esq. 120, 2do Piso. Tel: +54 221 4223528
jdiaz@unlp.edu.ar , ales@info.unlp.edu.ar , pamadeo@linti.unlp.edu.ar, iharari@info.unlp.edu.ar,
aosorio@cespi.unlp.edu.ar

RESUMEN

Los conceptos de usabilidad y accesibilidad pretenden mejorar la experiencia del usuario y son una parte fundamental del diseño inclusivo, que realmente beneficia a todos los usuarios. La accesibilidad a la información digital es crucial para la inclusión social y la igualdad de oportunidades.

El estado del arte de la Inteligencia Artificial (IA) para el diseño de la experiencia de usuario (UX) evoluciona constantemente, con un enfoque cada vez mayor en el diseño de sistemas de IA transparentes, explicables y justos que den prioridad al control y la autonomía del usuario, protejan la privacidad y la seguridad de los datos de los usuarios y promuevan la diversidad y la inclusión en el proceso de diseño. Algunas herramientas de accesibilidad utilizan la Inteligencia Artificial (IA) para identificar y corregir automáticamente errores de accesibilidad en el código del sitio web. Sin embargo, a pesar de las ventajas previstas y del rápido crecimiento del mercado, estas herramientas se enfrentan a críticas significativas en la comunidad de la accesibilidad debido a la preocupación sobre su eficacia y sobre si satisfacen las necesidades del grupo de usuarios al que van dirigidas. Este artículo presenta el estudio de las técnicas y herramientas basadas en IA y que se utilizan en UX y en Accesibilidad para ser utilizadas en diferentes dominios y con diferentes destinatarios, entendiendo cómo se pueden aprovechar estas herramientas, lo que tiene importantes implicaciones prácticas y de investigación.

Palabras clave: accesibilidad, usabilidad, inteligencia artificial

CONTEXTO

Dentro de la Facultad de Informática de la Universidad Nacional de La Plata, en el LINTI, Laboratorio de Investigación en Nuevas Tecnologías Informáticas, desde hace tiempo venimos con un trabajo sostenido en torno a la accesibilidad que ha permitido construir una perspectiva que coloca la inclusión como un paradigma que significa construir oportunidades para todos y todas [1].

En la Facultad de Informática se utilizan plataformas de código abierto para las gestiones académicas desde hace más de quince años, incluyendo sistemas

de gestión de aprendizaje como Moodle, repositorios abiertos y sistemas de gestión administrativa, como SIU Guaraní. También se realizan continuos mecanismos de integración entre las diferentes plataformas, lo que permite una mayor flexibilidad y aprovechamiento en el uso de las mismas. Estas integraciones favorecen la usabilidad, ya que se trata de que tengan una interfaz uniforme y que el acceso sea vinculado entre ellas.

Las líneas de investigación que se mencionan en este artículo se desarrollan en LINTI de la UNLP y están enmarcadas en el proyecto I+D 11/F028 "De la Sociedad del Conocimiento a la Sociedad 5.0: un abordaje tecnológico y ético en nuestra región" que se desarrolló hasta el año 2024 y continúan en el proyecto "Tecnologías digitales para la Inclusión, la Equidad y la Sostenibilidad" a partir del 2015. Ambos proyectos se encuentran bajo la dirección del Lic. Javier Díaz.

Dada su relevancia, el tema de accesibilidad web se viene trabajando en la Facultad desde el año 2002, y se incorporó esta temática en el plan de estudios de las carreras que se dictan en la institución, a través de la asignatura Diseño Centrado en el Usuario. El libro "Guía de recomendaciones para diseño de software centrado en el usuario" registra aproximadamente 10.000 descargas desde el año 2013 [2]. También se institucionalizó su abordaje mediante la creación de una Dirección de Accesibilidad desde el año 2010, se desarrollan tesinas, trabajos de cátedra al respecto, proyectos de innovación y desarrollo con estudiantes de Informática y diferentes proyectos de extensión acreditados por la Universidad Nacional de La Plata.

1. INTRODUCCIÓN

La usabilidad y la accesibilidad son dos conceptos fundamentales de los sistemas interactivos. La usabilidad es «la capacidad de un software para ser comprendido, aprendido, utilizado y atractivo para el usuario, en condiciones específicas de uso» [3]. Los principales factores que miden la usabilidad son la navegabilidad, la facilidad de aprendizaje, la velocidad de uso, la eficiencia del usuario y la tasa de errores. La accesibilidad es una práctica que «garantiza que los sitios web, las tecnologías y las herramientas están diseñados para ser utilizados por personas con discapacidad» [4]. La accesibilidad web afecta a muchos tipos de discapacidades, como las visuales, auditivas, físicas, cognitivas, neurológicas y del habla. Además, beneficia a otras personas, incluidas las mayores cuyas capacidades funcionales disminuyen como consecuencia del envejecimiento. La accesibilidad se mide a partir de herramientas de evaluación automática, evaluación manual por un experto y pruebas de usuarios con discapacidades [5].

La usabilidad es un concepto estrechamente relacionado con la accesibilidad. La usabilidad es una condición necesaria pero no suficiente para una buena accesibilidad. Y la accesibilidad por sí sola no garantiza la usabilidad [6]. Por ejemplo, un sistema puede ser utilizable para algunos usuarios. Sin embargo, puede no ser accesible para personas con discapacidad porque, por ejemplo, no ofrece un buen contraste de colores y los usuarios con baja visión no pueden percibir los elementos con los que pueden interactuar. Una aplicación accesible, que cumple los criterios de accesibilidad, no garantiza una buena experiencia de uso para todos los usuarios. A pesar de ser técnicamente accesible, puede tener una estructura de

navegación confusa. Por tanto, diseñar un sistema que cumpla los objetivos de usabilidad y accesibilidad pasa por tratar ambos aspectos [7].

La accesibilidad está vinculada con la usabilidad, ya que ambas son partes fundamentales en el diseño de la experiencia de usuario. La usabilidad debería incluir aspectos de accesibilidad, dado que un producto que no es accesible tampoco es usable para las personas con discapacidad. Sin embargo, en la práctica el concepto de usabilidad no tiende a enfocarse específicamente en la experiencia del usuario con alguna discapacidad. La accesibilidad, en cambio, se refiere a si todos los usuarios pueden acceder a una experiencia equivalente independientemente de cómo encuentran un producto o servicio [8].

La IA se presenta como una nueva tendencia con profundas implicaciones para la sociedad, que ofrece ventajas considerables a las personas con discapacidad. Sin embargo, junto a su promesa, la IA presenta riesgos inherentes, que podrían aumentar la discriminación contra las comunidades marginadas [9]. Por otro lado, introducir nuevas herramientas de IA en el proceso de diseño de la experiencia del usuario tiene el potencial de mejorar la eficiencia y la precisión, al tiempo que crea soluciones más innovadoras y creativas, pero aún son escasos los trabajos sobre cómo puede introducirse en procesos que dependen en gran medida de la creatividad y la aportación humanas [10].

En las próximas secciones se desarrollarán las líneas de investigación que se llevan a cabo sobre este tema, haciendo una revisión sistemática de la literatura para analizar cómo se utiliza la inteligencia artificial en el diseño de UX y en la evaluación y corrección de los aspectos de accesibilidad.

2. LÍNEAS DE INVESTIGACIÓN, DESARROLLO E INNOVACIÓN

Siguiendo con la línea de trabajo que venimos llevando a cabo desde hace varios años, en relación a la accesibilidad, como se plantea en [11] y vinculándola con la usabilidad de los sistemas y aplicaciones, las tareas que desarrollamos comprenden el estudio constante de nuevas normativas, herramientas emergentes, que permiten la adaptación de las plataformas y el desarrollo de aplicaciones destinadas a personas con algún tipo de discapacidad.

En estas líneas de investigación, las tareas que se llevan a cabo son las siguientes:

- Estudiar los métodos cualitativos que se aplican para evaluar la conformidad de un sitio web con las WCAG (Pautas de Accesibilidad al Contenido en la Web) mediante el uso de herramientas automáticas [12]
- Evaluar algunas heurísticas realizadas manualmente por expertos para el testeo de la accesibilidad de los sitios y aplicaciones web [13]
- Investigar sobre la accesibilidad en documentos, con especial atención en el impacto que tiene sobre grupo de personas con discapacidad visual y ceguera, en un ambiente de enseñanza.
- Establecer una metodología uniforme para el desarrollo de aplicaciones donde se consideren cuestiones de usabilidad y su repercusión en la accesibilidad de las mismas.

- Analizar interfaces adecuadas de los sistemas y aplicaciones que faciliten su utilización para todas las personas por igual.
- Estudiar cómo se utiliza la IA en el diseño de UX en la actualidad. Por ejemplo, para comprender el contexto de uso, descubrir las necesidades de los usuarios, ayudar a diseñar soluciones, evaluar el diseño y contribuir al desarrollo de soluciones.
- Estudiar las herramientas existentes que utilizan IA en el campo de la accesibilidad, tanto para testeos, como para la corrección de posibles errores.
- Estudiar y analizar aplicaciones orientadas a la ayuda e inserción de personas con alguna discapacidad.

Según el eje temático que venimos llevando respecto al estudio de la accesibilidad en todos los aspectos, es importante mencionar que desde diferentes espacios académicos e institucionales desplegamos prácticas educativas desde esta perspectiva. Todos los procesos de formación que se vienen impartiendo tienen como eje el reconocer la relevancia de trabajar la inclusión y accesibilidad, entendiendo que la enseñanza es la herramienta para plantear nuevos problemas que desafían la formación profesional.

3. RESULTADOS Y OBJETIVOS

Estas actividades ininterrumpidas que impulsan el reconocimiento y tratamiento de la accesibilidad en forma institucional, van generando conciencia y compromiso en todos los actores dentro de la Facultad.

Los resultados que se vienen logrando a través los años de estudio y trabajo se ven reflejados en los desarrollos y capacitaciones que se realizan en la comunidad académica y en la sociedad en su conjunto. El propósito general es evaluar la usabilidad y accesibilidad conjuntamente. Los objetivos que se plantean en este período son:

- Interactuar con organizaciones no gubernamentales, asociaciones civiles, centros de ayuda con el objetivo de interpelar sobre posibles soluciones que puedan dar respuesta a sus necesidades y requerimientos, con el objetivo de una mejor interacción
- Analizar los aspectos interrelacionados de usabilidad y accesibilidad de los sistemas y aplicaciones que se utilizan, en el ámbito de la Facultad.
- Analizar la accesibilidad en documentos respecto de los aspectos disciplinares y su composición respecto a elementos tales como textos, imágenes, y distintas estructuras.
- Continuar y mejorar la herramienta desarrollada que permite, en forma automática, solucionar problemas de accesibilidad en documentos de lectura, atendiendo las necesidades de las personas con discapacidad visual y ceguera.
- Continuar utilizando técnicas de IA para mejorar la automatización en cuestiones de accesibilidad, entre ellas las que abordan las funciones de reconocimiento de voz para conversión a texto, procesamiento del lenguaje

natural, texto predictivo, generación automática de textos alternativos para las imágenes.

Los resultados de las actividades que se vienen realizando son muy alentadores y satisfactorios, donde se demuestra que la accesibilidad web no es un concepto teórico y utópico. Los participantes comienzan a tomar conciencia de las barreras existentes en el acceso a las tecnologías, desarrollando sentimientos de empatía al ver experiencias concretas de gente con algún tipo de discapacidad.

La investigación futura también puede incluir el desarrollo de IA para la supervisión continua y automática y herramientas de ajuste. Estas herramientas permitirían una evaluación continua de la accesibilidad de los sitios web, lo que permitiría la identificación en tiempo real de problemas de usabilidad y sugerencias de mejora de conformidad.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo se encuentra formado por tres profesoras de amplia trayectoria en el campo de la investigación, que trabajan en el área de ambientes virtuales de aprendizaje y accesibilidad web. Además, una alumna becaria de la carrera de Licenciatura en Informática y un alumno tesista de la carrera de Licenciatura en Sistemas de la Facultad de Informática.

La participación en eventos de la especialidad, ha permitido seguir estableciendo canales de comunicación con otros investigadores que trabajan en las mismas áreas.

En la Facultad de Informática se realizan cada año las *Hackaton por la Inclusión*, que reúne a alumnos con el objetivo de trabajar en conjunto, en pos de lograr desarrollos accesibles orientados a la comunidad.

Las Jornadas de Accesibilidad que se realizan todos los años permiten intercambiar experiencias entre los participantes del curso y exponer los trabajos realizados.

5. REFERENCIAS

[1] Tauber, F. (2020). El desafío de gestionar en un escenario de pandemia. Disponible en: http://sedici.unlp.edu.ar/handle/10915/108475

[2] Díaz, J., Amadeo, P., Harari, I. "Guía de recomendaciones para diseño de software centrado en el usuario". Editorial de la Universidad Nacional de La Plata (EDULP). ISBN: 978-950-34-1030-1. Disponible en: http://sedici.unlp.edu.ar/handle/10915/32172

[3] ISO/CD 9241-11; Ergonomics of Human-System Interaction–Part 11: Guidance on Usability. ISO: Geneva, Switzerland, 1998. Disponible en: http://www.iso.org/iso/catalogue_detail.htm?csnumber=63500.

[4] Introduction to Web Accessibility. 2018. Disponible en: https://www.w3.org/standards/webdesign/accessibility

[5] Evaluating Web Accessibility Overview. 2020. Disponible en: https://www.w3.org/WAI/test-evaluate/

[6] Quintal, C.; Macías, J.A. Measuring and improving the quality of development processes based on usability and accessibility.

Univ. Access Inf. Soc. 2021, 20, 203–221.

[7] Pascual Almenara, A.; Sayago Barrantes, S. Towards Designing a Set of Usability and Accessibility Heuristics Focused on Cognitive Diversity: An Exploratory Case Study with Generative Artificial Intelligence. Information 2024, 15, 769. https://doi.org/10.3390/info15120769

[8] Interaction Design Foundation - IxDF. (2016, June 1). What is Accessibility?. Interaction Design Foundation - IxDF. https://www.interaction-design.org/literature/topics/accessibility

[9] Shuford , J. (2023). Contribution of Artificial Intelligence in Improving Accessibility for Individuals with Disabilities. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 2(2), 421-433. https://doi.org/10.60087/jklst.vol2.n2.p433

[10] Stige, Å., Zamani, E.D., Mikalef, P. and Zhu, Y. (2024), "Artificial intelligence (AI) for user experience (UX) design: a systematic literature review and future research agenda", Information Technology & People, Vol. 37 No. 6, pp. 2324-2352. https://doi.org/10.1108/ITP-07-2022-0519

[11] Díaz J., Harari I., Schiavoni A., Amadeo A., Osorio, A.: “Accesibilidad en diversos dominios de aplicación”. Publicado en el Libro de Actas del XXVI Workshop de Investigadores en Ciencias de la Computación - WICC 2024. 1a ed. - Comodoro Rivadavia: Universitaria de la Patagonia -EDUPA, 2024. Abril 18-19, Puerto Madryn, Chubut, Argentina. ISBN 978-987-8352-57-2

[12] Web Accessibility Evaluation Tools List. 2024. Available online: https://www.w3.org/WAI/test-evaluate/tools/list/

[13] Brajnik, G.; Vigo, M. AutomaticWeb Accessibility Metrics: WhereWeWere and WhereWeWent. In Web Accessibility; Human–Computer Interaction Series; Springer: London, UK, 2019.

Evaluación de la Experiencia del Usuario en Plataformas de Aprendizaje Virtual mediante Reconocimiento de Emociones

L. Aballay[1], I. Millán Tejada[2], S. Aciar[1], L. Romera[2], L. Espinosa[2], L. Nanni[1]
lnaballay@gmail.com, flavia.millan@gmail.com, saciar@gmail.com, lilianaromera@gmail.com, lvilmaespinosa@gmail.com, nannimatias@gmail.com

[1] Instituto de Informática, Facultad de Ciencias Exactas Físicas y Naturales, Universidad Nacional de San Juan
[2] Departamento de Informática, Facultad de Ciencias Exactas Físicas y Naturales, Universidad Nacional de San Juan

RESUMEN

Este artículo presenta el avance de un proyecto de investigación cuyo objetivo es diseñar una metodología para evaluar la experiencia del usuario (UX) en sistemas interactivos, específicamente plataformas de aprendizaje virtual, utilizando métodos de reconocimiento de emociones. En los objetivos del proyecto, se contempla una revisión bibliográfica, para construir un estado del arte preliminar, permitiendo establecer las bases de la metodología. Se presenta, el progreso en el primer año, y los hallazgos sobre los métodos de reconocimiento de emociones. Además, de identificar tendencias y patrones en la detección de emociones, con aplicación en plataformas de aprendizaje virtual en español. Los resultados preliminares sugieren que la combinación de métodos implícitos y explícitos de reconocimiento de emociones puede mejorar la evaluación de UX en entornos educativos digitales. Entre los resultados obtenidos, se mencionan las publicaciones en congresos nacionales, internacionales y revistas del área. Como aporte significativo, es importante mencionar que al momento de implementar la metodología de evaluación UX, se puede implementar en un sistema de cualquier ámbito: transporte, salud, educación, seguridad, marketing, entre otros. Esto le llevaría a cualquier institución que la aplique, evaluar la percepción de los usuarios, convirtiéndose en un aporte significativo para los diseñadores de estos sistemas.

Palabras claves: Computación Afectiva, Métodos de Evaluación de UX, Plataformas de Aprendizaje Virtual

CONTEXTO

El presente trabajo forma parte de la propuesta de investigación dentro del marco de un proyecto postulado a la convocatoria 2023-2025 del Consejo de Investigaciones Científicas y Técnicas y de Creación Artística (CICITCA) de la Universidad Nacional de San Juan (UNSJ). Las actividades de investigación se llevan a cabo en las instalaciones del Instituto de Informática (IdeI) de la Facultad de Ciencias Exactas físicas y Naturales (FCEFN) de la UNSJ, en el Laboratorio de Tecnologías Adaptativas y Colaborativas (TAC).

1. INTRODUCCIÓN

La experiencia del usuario (UX) es un factor crítico en el éxito de los sistemas interactivos, especialmente en plataformas de aprendizaje virtual, donde la satisfacción del usuario influye directamente en la motivación y el rendimiento académico [1] [2]. La evaluación de UX tradicionalmente se ha centrado en métricas de usabilidad y satisfacción, pero recientemente se ha reconocido la importancia de incorporar el análisis emocional para comprender mejor las respuestas de los usuarios [3] [4]. Este proyecto busca contribuir al campo de la educación virtual en español mediante el diseño de una metodología que integre métodos de reconocimiento de emociones para evaluar UX en sistemas interactivos.

1.1.Objetivos del Proyecto

Los objetivos del proyecto de investigación se detallan a continuación:

Objetivo General:
Diseñar una metodología para evaluar UX en sistemas interactivos mediante reconocimiento de emociones.

Objetivos Específicos:

- ✓ Identificar parámetros de evaluación en sistemas interactivos.
- ✓ Investigar métodos de reconocimiento de emociones y evaluación de UX.
- ✓ Diseñar la metodología de evaluación de UX con reconocimiento de emociones.
- ✓ Validar la metodología en un caso de estudio.

1.2. Metodología

Para abordar el proyecto, a través de los objetivos anteriormente planteados, es importante evaluar y comprender las emociones y respuestas emocionales de los usuarios al interactuar con estas plataformas. Teniendo en cuenta las emociones de los usuarios, los diseñadores pueden mejorar el diseño y la eficacia de las plataformas de aprendizaje virtual [5]. Al respecto, se desarrollaron los siguientes recorridos, para cumplir con las fases del cronograma de actividades propuesto. Siguiendo lo anterior, durante el primer año de ejecución, se han completado las siguientes fases:

- **Identificación de parámetros a evaluar en sistemas interactivos:** Se realizó una revisión bibliográfica exhaustiva para identificar las funcionalidades y parámetros clave que deben evaluarse en sistemas interactivos. Esta fase se completó al 100%, identificando métricas como usabilidad, accesibilidad, satisfacción entre otras.

- **Investigación sobre métodos de reconocimiento de emociones:** Se llevó a cabo una revisión bibliográfica sobre los tipos y métodos de reconocimiento de emociones, incluyendo técnicas como el análisis de expresiones faciales [6] [7] [8] [9], reconocimiento de voz [10] [11] y mediciones fisiológicas [12] [13]. Además, se construyó un estado del arte preliminar, completado al 70%.

Como resultado de estas actividades desarrolladas, se obtuvieron publicaciones en revistas del área y publicaciones y exposiciones en congresos y eventos locales, nacionales e internacionales, por parte de los miembros del proyecto, logrando establecer el estado del arte para poder continuar.

1.3. Métodos de Reconocimiento de Emociones

Se exploraron métodos de reconocimiento de emociones implícitos (automáticos) [14], como:

- Análisis de expresiones faciales.
- Reconocimiento de voz.
- Mediciones fisiológicas (frecuencia cardíaca, conductancia de la piel).
- Análisis de texto.

Se analizó la precisión, el tiempo de procesamiento, el coste, la intrusividad y la necesidad de un entorno de laboratorio para cada método.

La siguiente tabla comparativa, destaca las ventajas distintivas de cada método de detección de emociones, permitiendo a los investigadores y diseñadores de plataformas de aprendizaje virtual en español tomar decisiones informadas sobre el enfoque más apropiado para evaluar la experiencia del usuario y mejorar la calidad del aprendizaje en línea en español.

	texto	facial	voz
Precisión	✓	✓	✓
Tiempo de procesamiento	-	✓	✓
No invasiva	✓	-	✓
Requiere Entorno de Laboratorio	✓	✓	-
Recursos	✓	✓	✓
Factibilidad de aplicar en VLE	✓	✓	✓

1.4. Opiniones de Expertos

Se realizaron encuestas a expertos en educación virtual, interacción persona-ordenador y análisis emocional. Los resultados destacan:

- La importancia de la detección de emociones para comprender la UX.
- La preferencia por métodos no invasivos.
- El potencial de combinar métodos de reconocimiento de emociones con cuestionarios.

1.5.Discusión

La detección de emociones es fundamental para comprender la experiencia del usuario en plataformas de aprendizaje virtual. Los métodos implícitos, como el análisis facial y vocal, ofrecen ventajas en términos de precisión y no invasividad, pero requieren entornos controlados. Por otro lado, los métodos explícitos, como cuestionarios, son menos intrusivos, pero dependen de la autopercepción del usuario. La combinación de ambos enfoques puede proporcionar una evaluación más completa y precisa de UX.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Este proyecto ha sentado las bases para el diseño de una metodología de evaluación de UX en sistemas interactivos mediante reconocimiento de emociones. Como trabajo futuro y proponiendo líneas de investigación y desarrollo, se planea diseñar y validar un cuestionario para evaluar emociones en plataformas de aprendizaje virtual en español. Así mismo, continuar con el desarrollo y validación de la metodología propuesta. Todo lo anteriormente planteado, es con el fin de enriquecer aún más la evaluación de la experiencia del usuario en entornos de aprendizaje virtual.

3. RESULTADOS OBTENIDOS

Con los resultados de la investigación se realizarán publicaciones en congresos nacionales e internacionales, así como también se publicarán resultados más significativos en reconocidas revistas del área. Por otra parte, al momento de implementar la metodología de evaluación UX, es posible elegir como caso de estudio un sistema de cualquier ámbito: transporte, salud, educación, seguridad, marketing, por nombrar algunos. Realizar esta evaluación de la percepción de los usuarios será un aporte significativo para los diseñadores de estos sistemas, ya sea que pertenezcan a organismos públicos como a empresas privadas de desarrollo de software. Una vez hechas las evaluaciones a los sistemas, se le puede brindar a la empresa u organismo propietario del sistema, un análisis de los resultados obtenidos para que puedan tomar sus medidas correctivas pertinentes en caso de ser necesarias.

Además, el proyecto está estrechamente vinculado con las carreras de informática de la FCEFN; ya que en la materia “Ingeniería de Software II”, de la Licenciatura en Sistemas y la Licenciatura en Computación, y “Diseño Web II”, de la Tecnicatura en Programación Web, tienen entre sus contenidos mínimos la temática HCI y Experiencia de Usuario, lo que permitirá hacer participar a alumnos como posibles sujetos en pruebas experimentales con la implementación de la metodología de evaluación UX diseñada.

4. FORMACIÓN DE RECURSOS HUMANOS

En el marco del desarrollo del proyecto propuesto, se planea convocar a estudiantes a iniciar sus trabajos de fin de carrera de las carreras de Licenciatura en Ciencias de la Computación o de la Licenciatura en Sistemas de Información de la FCEFN de la UNSJ. En este sentido, se planea también postularlos en programas de becas de la UNSJ o Nación para alumnos avanzados. Por otro lado, también se propone convocar a alumnos de posgrado de la carrera de Maestría en Informática de la FCEFN de la UNSJ, para que realicen sus tesis en temas relacionados al proyecto. Adicionalmente, durante este proyecto se desarrollará una tesis de doctorado dentro del

programa de doctorado en Ciencias de la Informática de la FCEFN de la UNSJ.

5. BIBLIOGRAFIA

[1] N. Phongphaew and A. Jiamsanguanwong, "The Usability Evaluation Concerning Emotional Responses of Users on Learning Management System," *Proc. 2016 6th Int. Work. Comput. Sci. Eng. (WCSE 2016)*, pp. 43–48, 2016.

[2] E. Martínez Caro, "E-Learning: Un Análisis Desde El Punto De Vista Del Alumno," *RIED. Rev. Iberoam. Educ. a Distancia*, vol. 11, no. 2, pp. 151–168, 2012.

[3] M. Hassenzahl, "The Thing and I: Understanding the Relationship Between User and Product," Springer, Cham, 2018, pp. 301–313.

[4] L. N. Aballay, S. V. Aciar, and C. A. Collazos, "Emotions for Virtual Learning Environments," *IEEE-RITA*, vol. 9, no. 3, 2021.

[5] L. Aballay, A. Orellana, S. Aciar, and C. Collazos, "Affective Evaluation: Proposal for Educational Platforms," in *Jornadas Internacionales de Campus Virtuales*.

[6] E. Ivanova and G. Borzunov, "Optimization of machine learning algorithm of emotion recognition in terms of human facial expressions," *Procedia Comput. Sci.*, vol. 169, pp. 244–248, Jan. 2020.

[7] M. R. Mahmood *et al.*, "Classification techniques' performance evaluation for facial expression recognition," *Artic. Indones. J. Electr. Eng. Comput. Sci.*, vol. 21, no. 2, pp. 1176–1184, 2021.

[8] S. Patil, A. Joshi, G. Deore, A. Taley, and S. Sawant, "Emotion Detection Using Facial Recognition," *SSRN Electron. J.*, 2020.

[9] A. Vartanov, V. Ivanov, and I. Vartanova, "Facial expressions and subjective assessments of emotions," *Cogn. Syst. Res.*, vol. 59, no. 2020, pp. 319–328, 2020.

[10] E. S. Salama, R. A. El-Khoribi, M. E. Shoman, and M. A. Wahby Shalaby, "A 3D-convolutional neural network framework with ensemble learning techniques for multi-modal emotion recognition," *Egypt. Informatics J.*, vol. 22, no. 2, pp. 167–176, Jul. 2021.

[11] L. Schoneveld, A. Othmani, and H. Abdelkawy, "Leveraging recent advances in deep learning for audio-Visual emotion recognition," *Pattern Recognit. Lett.*, vol. 146, pp. 1–7, Jun. 2021.

[12] S. Cano, L. García, C. A. Collazos, and S. Albiol, "Towards the Design of a System to Measure Emotions using Physiological Signals as Biofeedback. Case Study in Higher Education Students," *REHAB 2019*, 2019.

[13] J. A. Domínguez-Jiménez, K. C. Campo-Landines, J. C. Martínez-Santos, E. J. Delahoz, and S. H. Contreras-Ortiz, "A machine learning model for emotion recognition from physiological signals," *Biomed. Signal Process. Control*, vol. 55, p. 101646, Jan. 2020.

[14] L. Aballay, C. Collazos, and S. Aciar, "Emotions evaluation on Virtual Learning Environments: Students perceptions," in *VII Jornadas Ibero-americanas de Interação Humano-Computador - HCI 2021*, 2021.

EVALUACION DEL SISTEMA DE GESTION DE LA CALIDAD EN EL AMBITO EDUCACTIVO

Jorge A. Silvera, Angel R. Barberis, Lorena E. del Moral, Daniel A. Arias Figueroa
Centro de Investigación y Desarrollo en Informática Aplicada (C.I.D.I.A.)
Facultad de Ciencias Exactas – Universidad Nacional de Salta – Salta - Argentina
jsilvera@unsa.edu.ar, barberis@cidia.unsa.edu.ar, daaf@cidia.unsa.edu.ar

RESUMEN

El presente trabajo presenta la evaluación de un Sistema de Gestión de Calidad (SGC) implementado en el Departamento de Informática de la Facultad de Ciencias Exactas de la Universidad Nacional de Salta (U.N.Sa), específicamente en la carrera de Licenciatura en Análisis de Sistemas.

En la actualidad, la noción de educación de calidad implica que la mayoría de las instituciones educativas consideren la satisfacción de sus estudiantes como uno de sus objetivos principales. Para lograrlo, es necesario cumplir con una serie de condiciones, requisitos y estándares de calidad.

En este sentido, es fundamental que cada organización evalúe sus procesos mediante un SGC, lo que les permitirá gestionar de manera eficiente sus recursos, implementar políticas de calidad y asegurar una mejora continua en su entorno organizacional.

Palabras claves:
Normas de Calidad ISO 9.001, Sistema de Gestión de Calidad (SGC), Licenciatura en Análisis de Sistemas, Calidad en Educación.

CONTEXTO

En el Departamento de Informática de la Facultad de Ciencias Exactas de la Universidad Nacional de Salta (UNSa), se implementó en el año 2018 un modelo de Sistema de Gestión de Calidad (SGC) basado en la normativa ISO/IEC 9001:2015. Esta iniciativa tuvo como objetivo involucrar a los agentes directivos, administrativos, docentes y estudiantes, reconociendo que un SGC garantiza la eficiencia de diversos procesos. En particular, se centra en el seguimiento del desempeño de los alumnos, un aspecto fundamental del proceso de enseñanza y aprendizaje que asegura una formación adecuada para los futuros profesionales de la Licenciatura en Análisis de Sistemas (LAS).

Este trabajo se enmarca dentro de la línea de investigación en Calidad, en la cual se viene trabajando desde hace tiempo, del Centro de Investigación y Desarrollo en Informática Aplicada (C.I.D.I.A), que opera en la Facultad de Ciencias Exactas de la UNSa desde junio de 2002. C.I.D.I.A. surgió como una evolución del Laboratorio de Investigación y Desarrollo en Tecnologías Informáticas (L.I.D.T.I), que funcionó desde 1997 y dependía del Departamento de Matemática de la misma facultad.

La carrera Licenciatura en Análisis de Sistemas, con resolución ministerial N°1891/22 - C.S 135/10, 262/12, 042/14, 426/17 - RESCD-DEXA 403/12, 001/10, 440/17, se implementó en la Facultad de Ciencias Exactas de la UNSa en el año1984, originalmente contemplaba un plan de estudios de cuatros años, con un título intermedio al cumplirse los primeros dos años y medio de la carrera. En el año 1997 se aprobó una modificación del plan de estudios, llevando su desarrollo a cinco años.

1. INTRODUCCION

El presente trabajo se aborda el estudio y evaluación de un Sistema de Gestión de Calidad (SGC) basado en la familia de Normas ISO/IEC 9000, específicamente en la versión 9001:2015, y su aplicación en instituciones educativas, con un enfoque particular en el ámbito universitario. Es fundamental que las

organizaciones comprendan que dependen de sus clientes y/o usuarios, por ello, deben identificar sus necesidades actuales y futuras, satisfacer sus requerimientos y esforzarse por superar sus expectativas. Además, cada organización debe optimizar los recursos disponibles para mejorar su eficiencia.

Desde un punto de vista etimológico, el concepto de calidad proviene del latín qualitis, que se refiere al conjunto de cualidades que constituyen la manera de ser de una persona o cosa. Es sinónimo de términos como cualidad, clase, aptitud, excelencia, categoría, casta, nobleza y superioridad, entre otros (Ardon N. y Jara, M., 2005). La calidad no es un concepto sencillo, y su significado puede variar según las condiciones históricas, culturales, entre otras.

Durante la etapa de revisión sistemática de la literatura, enfocada en el tema calidad en el ámbito de la educción superior, vimos muchas definiciones de calidad, nosotros adoptamos la de la Real Academia Española (2006), se ofrecen las siguientes definiciones de calidad:

- Propiedad o conjunto de propiedades inherentes a una cosa, que permiten apreciarla como igual, mejor o peor que las restantes de su especie.
- En sentido absoluto, buena calidad, superioridad o excelencia.

Es conocido, en el ámbito del Departamento de Informática de la Facultad de Ciencias Exactas, la problemática de baja cantidad de graduados en los últimos años en la carrera Licenciatura en Análisis de Sistemas. De los datos obtenidos en la Dirección de Alumnos podemos observar lo siguiente:

Año	**Egresados**
2011	26
2012	13
2013	20
2014	31
2015	3
2016	17
2017	8
2018	6
2019	12
Total	136

Figura 1 - Egresados de LAS, por año.

Tanto los alumnos como las autoridades reconocen la existencia de problemas significativos, como la falta de presupuesto, la escasez de bibliografía y equipamiento informático para la enseñanza, así como laboratorios obsoletos. Además, se cuenta con un número insuficiente de docentes, una infraestructura inadecuada y una total carencia de reconocimiento e incentivos para el personal docente y de apoyo universitario. Todo esto impacta de manera directa en la insatisfacción general de los alumnos.

En este contexto, los estudios sobre la satisfacción estudiantil en las universidades son fundamentales como indicadores para evaluar la calidad educativa. Una mayor satisfacción estudiantil está relacionada con un mejor rendimiento académico, lo que a su vez facilita la graduación (Garbanzo, 2017).

En Argentina, la Comisión Nacional de Evaluación y Acreditación Universitaria (CONEAU) comenzó a funcionar en 1996, realizando grandes esfuerzos para garantizar una educación de calidad a través de auditorías que promueven la mejora continua en la calidad educativa, siguiendo un proceso de autoevaluación (Flores, J. & Nicolosi, A., 2009).

Una alternativa adecuada para abordar la problemática observada es la implementación exitosa de un modelo de Sistema de Gestión de Calidad (SGC) que enfoque los procesos de la organización y oriente sus resultados hacia el

alumno, garantizando un nivel adecuado de eficacia.

La familia de Normas ISO/IEC 9000, desarrollada por la Organización Internacional de Normalización (ISO) en colaboración con la Comisión Electrotécnica Internacional (IEC), establece un modelo que contempla aspectos clave para una gestión de calidad basada en procesos. Este enfoque asegura la calidad de productos, ya sean tangibles o intangibles, y aumenta la satisfacción del cliente, que en este caso son los estudiantes.

En el Departamento de Informática de la Facultad de Ciencias Exactas de la Universidad Nacional de Salta, se implementó un modelo de SGC en 2018 en la carrera de Licenciatura en Análisis de Sistemas.

La calidad en educación se considera un aspecto fundamental que se asocia con la excelencia y se puede entender desde dos concepciones principales. La primera se basa en los recursos con los que cuenta la institución, tales como financiación, personal docente y de investigación de alta calidad, así como estudiantes de alto rendimiento, para los cuales existen criterios de evaluación específicos. La segunda concepción está relacionada con la reputación adquirida por la institución, bajo la premisa de que aquellas con mayor excelencia académica disfrutan de una mejor reputación.

Las diversas concepciones de la calidad en la educación superior y su relevancia han sido presentadas por Harvey y Green (1995), adaptadas al contexto latinoamericano por Mora Ruiz (1999). Ruiz explica que en la educación superior es posible diferenciar, al menos, seis formas distintas de interpretar el concepto de calidad:

- La calidad como producción original.
- La calidad como resultado de la correlación entre metas, objetivos y resultados.
- La calidad como la mayor o menor distancia con respecto a un modelo que se considera como lo ideal o deseable para la institución o unidad académica.
- La calidad como satisfacción de las necesidades del usuario.
- La calidad como valor agregado.
- La calidad como un agregado de propiedades.

El objetivo es investigar en qué medida la implementación de un SGC eficaz que cumpla con los requisitos de la familia de Normas ISO/IEC 9000 incide en la satisfacción de los alumnos y del personal de la UNSa, especialmente en la carrera de LAS.

Promover la calidad en la enseñanza universitaria es tratar de mejorar de forma continua sus prácticas de forma que permitan:

- Garantizar los resultados y productos obtenidos.
- Asegurar la trazabilidad de los procesos que se ejecutan en la organización.

Desde el momento de su creación, los requisitos de la norma internacional ISO 9.001:2015 son genéricos y aplicables a todas las organizaciones sin importar su tipo, tamaño o producto o servicio suministrado.

Podemos definir una metodología para implementar el SGC en el área de la capacitación, utilizado el esquema anterior de manera sistemática, dando lugar a un ciclo de mejora continua, tal y como se representa en la figura 2.

Otra contribución significativa de la Familia de Normas ISO/IEC 9000 ha sido la aplicación de los conceptos de "planear, hacer, verificar, actuar" en la calidad de la educación. Esta metodología, propuesta por Deming (1986), se presenta como un círculo virtuoso de mejora continua.

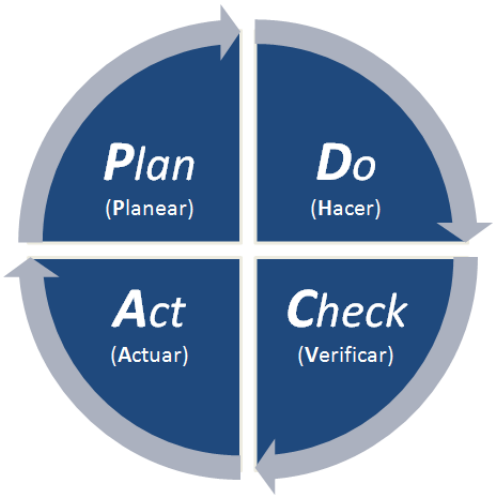


Figura 2- Ciclo de Deming – PDCA de mejora continua.

El trabajo realizado tuvo un alcance correlacional causal, ya que se analizó la relación entre los principios de un Sistema de Gestión de la Calidad (SGC) basado en la norma ISO/IEC 9001:2015 y su influencia en los procesos de capacitación y en la satisfacción del alumno. Se crearon dos instrumentos, validados por el grupo de investigación en calidad del C.I.D.I.A. El primer cuestionario se utilizó para medir la variable gestión de la calidad, constando de 12 indicadores con 5 escalas. El segundo cuestionario, enfocado en la satisfacción de los alumnos, también incluyó 12 indicadores, pero con 4 escalas. Para la recolección de datos se empleó la técnica de la encuesta, dirigida a la población mencionada anteriormente.

Para determinar el grado de confiabilidad, se llevó a cabo un análisis de consistencia interna utilizando el coeficiente alfa de cronbach, que permite medir las correlaciones entre las variables y evaluar la confiabilidad o homogeneidad de las preguntas en la escala de Likert, definiendo las siguientes variables:

- Variable independiente: Sistema de Gestión de la Calidad.
- Variable dependiente: Satisfacción del alumno.

Se utilizo el método deductivo inductivo, se utilizaron las siguientes técnicas de investigación:

- Entrevistas: Durante el desarrollo de la investigación, se realizaron entrevistas dirigidas a alumnos y personal de apoyo universitario, con el objetivo de comprender la problemática en estudio.
- Análisis documental: Se revisó la documentación de la cátedra, la carrera, el departamento de informática y la dirección de alumnos, verificando la situación en relación con los requerimientos de la norma.

La población muestral está compuesta por alumnos de la carrera de Licenciatura en Análisis de Sistemas (LAS), correspondientes a los años 2018 y 2019.

Interpretamos que los resultados de la prueba de hipótesis obtenidos se encuentran bastante lejos del valor crítico para rechazar la hipótesis nula, ya que 0.93 es menor que 1.67. Esto sugiere que no hay suficiente evidencia para rechazar la hipótesis nula en este caso.

3. LINEAS DE INVESTIGACION y DESARROLLO

Los principales ejes temáticos que se están investigando son los siguientes:

- Tecnología Informática aplicada en Educación.
- Gestión de Calidad aplicada a la educación superior.
- Análisis de Sistemas.
- Norma ISO/IEC 9.001.

3. RESULTADOS OBTENIDOS

A partir de la experiencia obtenida en la realización del presente trabajo, se puede concluir que la satisfacción de los estudiantes con la educación universitaria fue notablemente alta. en casi todas las

dimensiones y desagregaciones de las variables analizadas. Las valoraciones fueron predominantemente positivas, reflejando un alto grado de satisfacción, siendo relevante destacar que la categoría "totalmente insatisfecho" no fue reportada en el estudio, y que la opción "poco satisfecho" no superó el 12%. Estos resultados sugieren que la calidad de la educación en la carrera estudiada es, en general, satisfactoria.

El desarrollo e implementación del Sistema de Gestión de Calidad se basó en la norma ISO 9001:2015, alineándose con los objetivos del programa académico. Este enfoque permitió establecer los estándares necesarios para diseñar adecuadamente el Sistema de Gestión de Calidad, mejorando la eficacia de los procesos de un 65% a un 80%.

En el futuro, se contempla la posibilidad de alcanzar la certificación del proceso principal, así como también la extensión de esta experiencia a otras carreras de grado.

Finalmente, los sistemas de autoevaluación, evaluación externa y acreditación de instituciones y programas universitarios son elementos fundamentales para certificar la calidad de los procesos en las instituciones de educación superior, funcionando como motores clave de la gestión de la calidad universitaria.

4. FORMACION DE RECURSOS HUMANOS

La estructura del equipo de investigación está compuesta por tres miembros: un director y dos docentes.

Uno de los integrantes posee una Especialización en Ingeniería de Software de la Universidad Nacional de La Plata, habiendo presentado como trabajo final el proyecto titulado “Sistema de Gestión de Calidad bajo Normas ISO”.

Otro miembro obtuvo el título de Licenciado en Análisis de Sistemas por la Universidad Nacional de Salta, con la tesis “Análisis y Diseño de un Sistema de Gestión de Calidad basado en la Norma ISO 9.001”.

Finalmente, el tercer integrante alcanzó el título de Magíster en Administración de Negocios, con un trabajo final titulado “Evaluación del sistema de gestión de la calidad implementado en el ámbito educativo del Departamento de Informática de la Facultad de Ciencias Exactas” en la Escuela de Negocios de la Universidad Católica de Salta.

5. BIBLIOGRAFIA

Norma ISO 9001:2015. Editorial IRAM.

Garbanzo Vargas, G. (2017). Factores asociados al rendimiento académico en estudiantes universitarios, una reflexión desde la calidad de la educación superior pública. Revista Educación.

Flores, J., Nicolosi, A. (2009). Informe sobre la educación superior en Iberoamérica 2009 - capítulo sobre la república argentina dentro del Proyecto Alfa ALFA Nro. Buenos Aires: Cinda-Alfa.

Ardon, N. & Jara, M. (2005). Sistema para el monitoreo de la calidad de la atención en Salud por indicadores en los hospitales de I, II y III nivel de atención. Bogotá.

Castellón Murcia, Roberto Aristides (2009). Aplicación de la norma ISO 9.001 en el proceso enseñanza aprendizaje en la educación superior, UFG Editores.

Harvey, L. y Green, D. (1993) Defining Quality. Assessment and Evaluation in Higher Education.

Mora Ruiz, J.G. (1991). Calidad y Rendimiento en las Instituciones Universitarias. Madrid.

FACTORES QUE IMPACTAN EN LA PERMANENCIA Y GRADUACIÓN DE ESTUDIANTES EN CARRERAS DE SISTEMAS: UN ESTUDIO EN UNIVERSIDAD CAECE

Lucía Rosario Malbernat, María Patricia Clemens, Juan David, Cristian Ferreyra

Departamento de Sistemas, Universidad CAECE

lmalbernat@ucaecemdp.edu.ar

RESUMEN

Este estudio analiza la deserción, el desgranamiento y la demora en graduaciones de las carreras de Licenciatura e Ingeniería en Sistemas de la Universidad CAECE, con sedes en CABA y Mar del Plata.

A través de la aplicación de metodologías cuali-cuantitativas, combinando métodos estadísticos, algoritmos de minería de datos y técnicas cualitativas, se analizaron los datos de las cohortes del último quinquenio con el objetivo de identificar los factores endógenos y exógenos con incidencia en la permanencia y graduación de los estudiantes.

Los hallazgos indican que aspectos como la situación laboral, el rendimiento académico y la carga horaria son determinantes en la continuidad académica. Asimismo, se identificaron asignaturas que presentan mayores desafíos para los estudiantes.

Para abordar esta problemática, se proponen estrategias institucionales que buscan mejorar la retención y reducir los tiempos de graduación. Estas iniciativas no solo optimizan la experiencia académica de los estudiantes, sino que también contribuyen a satisfacer sus expectativas y las de sus familias. Además, responden a la creciente demanda de profesionales en el sector informático, un área clave para el desarrollo tecnológico y económico del país.

La investigación ha aportado una base de conocimiento que permitió diseñar políticas educativas más efectivas y adaptadas a las necesidades del estudiantado del Departamento de Sistemas de la Universidad CAECE.

***Palabras clave*:** Desgranamiento; deserción; sistemas; licenciatura; ingeniería.

CONTEXTO

Esta investigación está radicada en el Departamento de Sistemas de la Universidad CAECE, Argentina, correspondiente al período 2023-2024, con prórroga hasta 2025. Su inicio fue el 1 de febrero de 2023 y su finalización está prevista para el 31 de diciembre de 2025. Se trata de un proyecto intersedes, con la participación de investigadores de la Sede Central y de la Subsede Mar del Plata, cuyo impacto alcanza a la gestión de las carreras de grado del Departamento, Licenciatura e Ingeniería en Sistemas. Está aprobada por R.R. Nº 384/22.

1. INTRODUCCIÓN

El término desgranamiento se refiere a las demoras en el avance de los estudiantes respecto al ritmo previsto en el diseño curricular, mientras que deserción implica el abandono definitivo de los estudios [1].

El desgranamiento y la deserción en carreras universitarias de informática representan desafíos significativos para las instituciones educativas, especialmente en un contexto donde la demanda de profesionales en tecnología supera la oferta disponible. En Argentina, diversos estudios [1], [5], [7] señalan que menos del 30% de los estudiantes que inician carreras informáticas se gradúan en el tiempo estipulado, y las tasas de deserción pueden superar el 50% en los primeros años.

En la Universidad CAECE, si bien la tasa de deserción es inferior a la media nacional, la problemática del desgranamiento y la deserción sigue presente, al igual que la demora en la graduación.

Este estudio ha tenido como objetivo analizar los factores que inciden en estos fenómenos y desarrollar estrategias para mejorar la permanencia y el éxito académico, estrategias que ya se encuentran en proceso de implementación.

Comprender estos fenómenos no solo permite diseñar intervenciones institucionales más efectivas, sino que también tiene un impacto directo en la trayectoria profesional de los estudiantes. La mejora en la permanencia y graduación favorece a una mejor inserción laboral, alineada con las demandas del mercado, contribuyendo al crecimiento del sector tecnológico en Argentina.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El proyecto se ha centrado en los siguientes ejes de investigación:

- **Duración real de las carreras:** Evaluación del tiempo efectivo que los estudiantes demoran en completar sus estudios en comparación con la duración teórica establecida.
- **Desgranamiento cuatrimestral y deserción por cohorte y sede:** Análisis de la disminución progresiva de estudiantes activos por cuatrimestre y la tasa de abandono en cada cohorte y sede.
- **Factores intrínsecos y extrínsecos:** Análisis de causas internas (endógenas) y externas (exógenas) que afectan la continuidad académica, tales como situación laboral, rendimiento académico, métodos de enseñanza y apoyo institucional.
- **Estrategias de intervención:** Diseño de acciones concretas para abordar los factores identificados, incluyendo tutorías académicas, flexibilización de horarios y apertura de recursadas a contraturno.

3. RESULTADOS OBTENIDOS

El análisis de los datos recopilados reveló que la situación laboral de los estudiantes es uno de los factores extrínsecos más relevantes en el desgranamiento y la deserción, así como la principal causa de demora en la graduación.

Muchos estudiantes acceden a empleos en empresas informáticas reconocidas desde el segundo o tercer año de la carrera, lo que dificulta el cumplimiento de la carga horaria y afecta su rendimiento académico. Además, los programas de intercambio cultural que permiten trabajar en el exterior durante el receso estival limitan la disponibilidad para rendir exámenes finales, retrasando la aprobación de materias.

Asimismo, se identificó que las asignaturas con altos índices de recursado presentan una mayor tasa de abandono, lo que sugiere la necesidad de revisar sus contenidos y métodos de enseñanza, así como ofrecer recursadas a contraturno.

En cuanto a la duración real de las carreras, menos del 20% de los estudiantes logran graduarse en el tiempo estipulado, mientras que en Argentina aproximadamente el 25% alcanza la titulación en el tiempo teórico estimado [4]. Esto resalta la urgencia de implementar estrategias que faciliten la finalización de los estudios en tiempo y forma.

Tal como señalan Schmidt et al. [2], el seguimiento continuo del rendimiento académico y la inclusión de actividades de refuerzo son clave para reducir la deserción. En consecuencia, se han propuesto iniciativas como tutorías personalizadas, seguimiento académico individualizado, mayor flexibilidad horaria y la implementación de recursadas a contraturno. Algunas de estas medidas comenzarán a aplicarse en el ciclo lectivo 2025, por lo que se espera analizar su impacto en futuras ediciones del Workshop. Además, la prórroga del proyecto permitirá dar seguimiento a estas acciones y evaluar su efectividad en la mejora de la trayectoria académica del estudiantado.

4. FORMACIÓN DE RECURSOS HUMANOS

El grupo de investigación está compuesto por profesores del Departamento de Sistemas y directores de carreras. Durante el primer trimestre del proyecto, se incentivó la participación de estudiantes como investigadores iniciales. Hasta la fecha, un estudiante ha completado su participación y ha obtenido el título de Ingeniero en Sistemas.

5. BIBLIOGRAFÍA

[1] Leone, L., Veizaga, K., Conforte, J., & Zanazzi, J. L. (2014). Modelos para explicar el desgranamiento en una carrera de Ingeniería. En 112º Simposio Argentino de Investigación Operativa (SIO 2014), 43 JAIIO (pp. 25-32). ISSN: 1850-2865.

[2] Schmidt, B., Boero, P., Mendez Vera, J. (2023) Factores que influyen en la deserción universitaria: El caso de una universidad estatal chilena. Revista Portuguesa de Educação, vol. 36, no. 1, e23002, 2023.

Universidade do Minho

[3] Lázaro Alvarez, N., Callejas, Z., & Griol, D. (2020). Factores que inciden en la deserción estudiantil en carreras de perfil Ingeniería Informática. Revista Fuentes, 22(1), 105–126. https://doi.org/10.12795/revistafuentes.2020.v22.i1.09

[4] Sanseau, M, Sanchez Cestona, J, Calio, S (2023) Permanencia de las y los estudiantes en la universidad 1a ed. - Ciudad Autónoma de Buenos Aires: CONEAU-Comisión Nacional de Evaluacion y Acreditación Universitaria, ISBN 978-987-3765-75-9

[5] Moreno, J.E, Chiecher, AC (2019) Abandono en carreras de Ingeniería. Un estudio de los aspectos académicos, socio-demográficos, laborales y vitales. Cuad. Investig. Educ. vol.10 no.2 Montevideo dic. 2019. ISSN 1510-2432; versión On-line ISSN 1688-9304. https://doi.org/10.18861/cied.2019.10.2.2908

[6] Lorenzi, G. (2024). Deserción en la educación superior Argentina. Las competencias digitales como nuevo desafío para la inclusión. Revista De Ciencia, Tecnología E Innovación, 22(30), 269–278. https://doi.org/10.56469/rcti.v22i30.968

[7] Minnaard, C, Torres, Z, Minnard, V (2024) Deserción y retención, indicadores de la calidad educativa en carreras de ingeniería desde la formación por competencias. XIX Congreso de Tecnología en Educación y Educación en Tecnología ISSN: 978-987-3977-95-4

Investigación y desarrollo de Tecnologías Digitales para Educación

Sanz Cecilia[1,3][0000-0002-9471-0008], Artola Verónica[1] [0000-0003-0833-1077], Gonzalez Alejandro[1][0000-0002-9854-8618], Zangara Alejandra[1][0000-0002-9854-8618], Iglesias Luciano[10000-0002-3284-6997], Ibáñez Eduardo[1] [0000-0002-1211-9370], Violini Lucía[1,2][0000-0002-0537-5528], Fachal Adriana[1][0000-0002-6379-2856], Del Gener, Aldana[1][0009-0008-0024-3703], Abásolo María José[1,3][0000-0003-4441-3264], Dieser Paula[4][0000-0002-2066-0183], Coma-Roselló Teresa [5][0000-0002-7439-8459] , Handula Lorenzo[1], Lombardo Sabrina[1], Segovia Maite[1], Pesado Patricia[1{0000-0003-0000-3482}]

[1]Instituto de Investigación en Informática LIDI (III-LIDI). Centro Asociado CIC.
Facultad de Informática – Universidad Nacional de La Plata

[2] Ha contado con beca UNLP

[3]Investigador Asociado de la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC)

[4] Departamento de Matemática, Facultad de Ciencias Exactas y Naturales, Universidad Nacional de La Pampa.

[5] Departamento de Ciencias de la Educación, Universidad de Zaragoza, España

{csanz, vartola, agonzalez, li, eibanez, lviolini, Aldanamdg@lidi.info.unlp.edu.ar, mjabasolo, ppesado}@lidi.info.unlp.edu.ar,
alejandra.zangara@gmail.com, afachal@hotmail.com, paula.dieser@gmail.com, sabrina.lombardo@alu.ing.unlp.edu.ar, lorenzohandula@gmail.com, tcoma@unizar.es

RESUMEN

Este trabajo presenta un subproyecto acreditado del Instituto de Investigación en Informática LIDI (III-LIDI), con algunas de sus principales líneas de investigación, desarrollo e innovación. En el subproyecto se estudian, a la luz de la Ingeniería de Software, el Diseño Instruccional, la Cognición humana, el Diseño de Interfaces y otras áreas, temáticas que se vinculan con la creación, evaluación e integración de tecnologías digitales en escenarios educativos. En particular, se detallan aquí estas líneas de investigación y los principales resultados alcanzados en el año 2024.

Palabras clave: materiales educativos digitales, entornos digitales para educación, juegos serios, herramientas colaborativas, educación, inteligencia artificial

CONTEXTO

El presente trabajo está vinculado al subproyecto de "Creación de tecnologías digitales para el escenario educativo" enmarcado en el proyecto: "Diseño, desarrollo y evaluación de sistemas en escenarios híbridos para áreas clave de la sociedad actual: educación, ciudades inteligentes y gobernanza digital" (2023-2026). Se trata de un proyecto del III-LIDI, de la Facultad de Informática de la Universidad Nacional de La Plata; acreditado por el Ministerio de Educación. Este proyecto da continuidad a las investigaciones y desarrollos que realiza el III-LIDI desde hace más de 20 años en el área de tecnología y educación.

1. INTRODUCCIÓN

A lo largo de la historia de la educación y atendiendo a criterios de calidad, las escuelas, universidades y centros de investigación han introducido innovaciones educativas: se han instaurado y diversificado las prácticas pedagógicas, tecnologías, estrategias educativas y didácticas, concepciones, modelos y métodos formativos, siempre en función de atender a prácticas y necesidades emergentes de la sociedad de cada momento (Rikkerink, et al., 2016). En consecuencia se espera que las instituciones educativas incluyan estrategias, tecnologías y experiencias de aprendizaje más proactivas y creativas que atiendan a las necesidades de la sociedad en cada contexto socio-cultural y político. En este proyecto se estudian metodologías, herramientas, y formas de uso de tecnologías digitales en procesos de enseñar y aprender, en función de las necesidades de la sociedad en general, y de los actores clave de cada institución.

Para abordar estos desafíos se entraman en este estudio saberes multi e interdisciplinares que iluminan la investigación y la creación de tecnologías digitales para el ámbito educativo. La co-creación de estas tecnologías con los miembros clave de la comunidad (Anderson-Coto, 2024) es una de los pilares para este subproyecto.

En los últimos años en este subproyecto se ha puesto el foco en líneas tales como: los entornos digitales y sus categorizaciones según sus posibilidades de mediación de situaciones educativas, el trabajo y aprendizaje colaborativo y su seguimiento con indicadores y herramientas innovadoras, la creación de materiales educativos digitales, la gamificación y los juegos serios y las posibilidades de la inteligencia artificial en el escenario educativo. Al mismo tiempo, un grupo de investigadores del proyecto se enfoca en estudiar modelos de interacción emergentes y sus potencialidades para crear experiencias educativas combinando diferentes sistemas, dispositivos, sensores y paradigmas de interacción (Gómez-Galán, 2020; Gros

y Durall, 2020). Esta última línea se presenta más detalladamente en otro trabajo.

1.1 Entornos digitales para mediación de situaciones educativas

En la actualidad existe diversidad de entornos digitales que han sido creados para mediar procesos educativos. Los reconocidos entornos virtuales de enseñanza y aprendizaje (EVEA) posibilitan la gestión y desarrollo de cursos centrados en la web con herramientas que facilitan la comunicación, la presentación de materiales de estudio, la producción conjunta, la evaluación, el seguimiento de los participantes, entre otros. También se utilizan entornos 3D orientados a la experimentación, y la simulación. Las tecnologías que subyacen a este tipo de entornos, su arquitectura, y funcionalidad son temas de estudio, así como su integración, y aportes en los procesos educativos.

Muchos de estos entornos proveen herramientas que soportan la colaboración, y el aprendizaje conjunto. Una tendencia en la agenda de investigación incluye el estudio de indicadores y posibilidades de reflejar la tarea de un grupo, mostrando su producción y sus intercambios. Otra de las tendencias es el estudio de las posibilidades y barreras que ofrece la inteligencia artificial en el trabajo con los EVEA.

1.2 Trabajo y aprendizaje colaborativo y su seguimiento con indicadores y herramientas innovadoras

Desde hace ya varios años que la mediación de proceso colaborativos con tecnologías digitales es un tema de investigación y desarrollo. Por ejemplo, el uso de herramientas tecnológicas digitales aporta en el aprendizaje individual de los actores educativos que intervienen en la colaboración. Pero también, ofrecen un gran potencial para promover la construcción conjunta del conocimiento, así como el desarrollo de habilidades relacionadas con la interacción que da lugar a procesos de aprendizaje más esenciales (Derouech et al., 2024). Las estrategias y herramientas de *mirroring* posibilitan una mayor conciencia del proceso colaborativo (van Leeuwen & Rummel, 2020).

1.3 La creación de materiales educativos digitales e inteligencia artificial

La creación de materiales digitales con fines educativos ha dado lugar a diferentes investigaciones y paradigmas de diseño. Las características de hipermedialidad, interactividad, granularidad y personalización vienen siendo estudiadas desde hace años. Los objetos de aprendizaje marcaron un hito en este campo al proponer materiales educativos de baja granularidad. La discusión sobre el nivel de granularidad y la posibilidad de generar itinerarios personalizados a partir de recomendaciones y ensamblado de objetos de aprendizaje sigue siendo un tema actual. Las nuevas posibilidades ofrecidas por las herramientas de inteligencia artificial generativa abren desafíos y necesidad de profundizar estudios en esta línea temática.

1.4 Gamificación y juegos serios

La gamificación aporta la introducción de elementos de los juegos en situaciones y actividades educativas. La creación de insignias, tablas de puntajes, rankings, son algunas de las mecánicas que hoy en día los docentes integran en sus propuestas. Se busca a través de estas posibilidades motivar e incrementar la participación activa de los estudiantes.

Los juegos serios educativos, en cambio, son juegos diseñados con un propósito que va más allá de entretener. Son utilizados como parte de las actividades para que el estudiante se apropie de saberes y/o habilidades.

Las líneas de I+D+i que se abordan se vinculan con las temáticas de la Maestría y Especialización en Tecnología Informática Aplicada en Educación de la misma Facultad. En estas carreras participan docentes-investigadores de este proyecto.

2. LÍNEAS DE INVESTIGACIÓN / DESARROLLO/ INNOVACIÓN

Se presentan aquí algunas de las principales líneas de investigación, desarrollo e innovación abordadas en el marco del proyecto:

- **Entornos digitales para la mediación de procesos educativos**: entornos virtuales de enseñanza y aprendizaje, entornos 3D, redes sociales, entornos virtuales gamificados. Funcionalidades y formas de acceso a estos espacios, trazabilidad de las actividades, estándares. Estrategias para el diseño e implementación de estos tipos de entornos. Potencialidades y barreras en la Integración de Inteligencia Artificial en estos entornos.
- **Materiales educativos digitales**. Metodologías para su diseño y producción. Objetos de aprendizaje. Multimedia e hipermedia en escenarios educativos. Nuevos entramados de medios, soportes y lenguajes. Estándares.
- **Integración de tecnologías digitales en procesos educativos**. Hibridación de las modalidades educativas. Diseño de MOOC. E-actividades. Actividades transmedia.
- **Gamificación y Juegos Serios Educativos con diferentes paradigmas de interacción.** Metodologías para su creación. Objetos activos.
- **Trabajo colaborativo mediado por tecnologías digitales**. Herramientas para promover la autorregulación y el desarrollo de capacidades metacognitivas. Conceptualización, análisis y desarrollo de software y metodologías. Actividades colaborativas aprovechando dispositivos móviles.
- **Formación de recursos humanos en el área de Informática.** Alfabetización digital, pensamiento computacional. Desarrollo de la empatía, y la

resiliencia académica en estudiantes de carreras en Informática.

3. RESULTADOS OBTENIDOS

En esta sección se describen los resultados alcanzados durante el año 2024.

Desde el año 2000 se ha trabajado en la investigación y desarrollo de entornos virtuales de enseñanza y aprendizaje. Se cuenta con un EVEA propio llamado IDEAS que ha sido desarrollado inicialmente en 2002 y que ha ido evolucionando (Sanz et al., 2006). En el 2024 se trabajó en una reestructuración de la forma en que se visualiza el itinerario de contenidos de un curso. Al mismo tiempo, se avanza sobre la posibilidad de publicar paquetes de contenidos estandarizados, y mejorar la creación de tareas para los estudiantes. Además, se tomaron nuevas decisiones en el pipeline de trabajo para optimizar las actualizaciones del sistema (Iglesias, 2024). Una de las integrantes del proyecto está abordando el estudio de herramientas de IA y llevó adelante una comparativa para su integración en IDEAS. Este trabajo forma parte de una práctica profesional supervisada finalizada de la carrera de Ingeniería en Computación (Lombardo & Sanz, 2024).

Además, se continúa avanzando en un trabajo de maestría en el que se realiza una revisión sistemática de literatura de herramientas de inteligencia artificial que podrían aportar al desarrollo de procesos educativos mediados por EVEA (García, Artola y Sanz, 2024).

En relación a la mediación de procesos educativos con EVEA, se finalizó una tesis referida a estrategias de aprendizaje en contextos rurales. Se trabajó con estudiantes del CEM Rural (Centro de Educación Media Rural) a través de entornos virtuales de enseñanza y aprendizaje de la provincia de Río Negro, donde se pudo encontrar que aprenden a autorregular su aprendizaje, a trabajar y aprender con otros/as, a gestionar estrategias de aprendizaje cooperativas y de autorregulación del propio proceso de aprendizaje (Honcharuk & Gonzalez, 2024).

También, se entregó una tesis doctoral para su evaluación y se realizaron publicaciones vinculadas a ésta sobre gamificación en EVEA. La tesis aporta un modelo tecno-pedagógico de gamificación en EVEA de utilidad para docentes e instituciones educativas. También presenta un estudio de casos donde se aplicó el modelo y se revelan los resultados y conclusiones (Vera et al., 2024).

Se avanzó, al mismo tiempo, en resultados relacionados con entornos virtuales 3D. Se finalizó una tesis doctoral relacionada con las posibilidades de estos entornos para procesos educativos orientados a personas con discapacidad auditiva. Uno de los aportes de la tesis consiste en una serie de 11 guías de buenas prácticas de diseño en este tipo de entornos y para esta población. También una metodología que propone etapas para desarrollar experiencias educativas siguiendo las guías propuestas. En la tesis se desarrollan experiencias concretas en OpenSIm, un motor de entornos 3D, donde se crearon escenarios ad-hoc. Las experiencias se desarrollaron en diferentes niveles educativos. (Fachal, Sanz y Abásolo, 2024; Fachal, Abásolo y Sanz, 2024).

Se ha trabajado en el desarrollo de hologramas y técnicas de mapeo de proyecciones que posibilitan experiencias educativas en entornos físicos aumentados e interactivos. De esta manera se crearon proyectos que resultan de estas investigaciones. Se detallan a continuación algunos de ellos.

Piano interactivo y proyecciones a partir de sus notas: se trata de un proyecto que se utiliza una interfaz de papel que simula ser un piano y a partir de tocar sus teclas se visualizan en el espacio físico, mapeando en objetos, diferentes personajes de videojuegos. Este proyecto permite recuperar la historia de los videojuegos y al mismo tiempo experimentar nuevas formas de interacción (Figura 1).



Figura 1 – Izquierda: estudiantes interactuando con el piano de papel; Derecha: instalación en el entorno donde se mapean proyecciones

Se continúa con la investigación en juegos serios educativos y exergames. En 2024 se finalizó un trabajo de especialización que aborda el tema de juegos serios educativos para matemática. Se presentó un trabajo que da cuenta de un estado del arte sobre esta temática (Gubaro, Sanz & Artola, 2024). En relación a los exergames se trabajó en el desarrollo de un dispositivo para controlar con las manos diferentes acciones en juegos destinados a promover la actividad física. En la Figura 2, derecha, se puede ver una imagen del prototipo que se está desarollando, basado en sensores (giroscopio, y de ultrasonido). Este nuevo prototipo se combinará con la tobillera ya desarrollada (Figura 2 izquierda) para sensar movimientos de caminar, saltar y correr en el lugar (Del Gener, Sanz e Iglesias, 2023).



Fig.2. Izquierda: tobillera para sensar movimientos de caminar, saltar y correr en el lugar. Derecha: Exergame Capitana Aldana y prototipo de sensor de movimiento de la mano (versión 2024)

Sobre la línea de materiales educativos, se finalizó el estudio de caso de la tesis doctoral sobre objetos de aprendizaje (OA) de uno de los miembros del proyecto (Violini, Sanz,y Pesado, 2023). En el estudio de caso se trabajó con docentes y estudiantes que utilizaron el framework creado para diseñar la propuesta de un OA. Se realizó también investigación teórica asociada a la temática de OA. La Figura 3 muestra una imagen de una de las pantallas de MarCOA, el *framework* desarrollado.

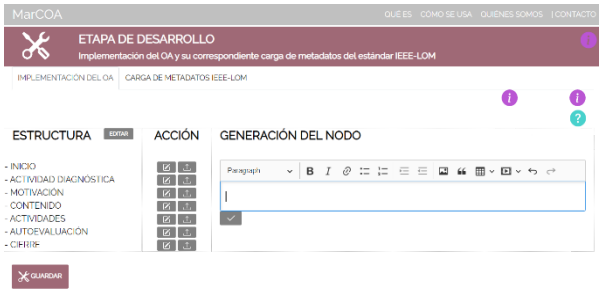


Fig. 3 – Pantalla de MarCOA: framework para el diseño de objetos de aprendizaje

Algunos de los principales resultados de las líneas vinculadas a trabajo colaborativo, metodologías y herramientas para su seguimiento, se enfocan en la investigación del estado del arte sobre *mirroring.* Se está finalizando una tesis doctoral en la que se creó *dashboard* para Moodle con el fin de implementar estrategias de *mirroring* (Vazquez et al., 2024).
También se avanzó en la aplicación de indicadores de seguimiento en procesos colaborativos en el marco del Seminario de Educación a Distancia de la Maestría en Tecnología Informática aplicada en Educación. Se trabajó en estrategias de *mirroring* que posibilitan corroborar resultados previos y generar nuevas hipótesis (Dieser, Sanz y Zangara, 2023a y b).
Se profundizó en investigaciones que se venían desarrollando en el marco de un proyecto de innovación docente de la Universidad de Zaragoza en el que se participa. En 2023 se trabajó en un modelo para el desarrollo de la empatía como competencia transversal en estudiantes de carreras en Informática (Sanz et al, 2023). Posteriormente, se trabajó en estrategias de gamificación que potencian estas temáticas. El juego EmpoderAR creado en el marco de este proyecto es utilizado en vinculación a estos temas y al acercamiento de jóvenes a las TIC y las ciencias en general (Sanz, Artola & Ibañez, 2024).
En 2024 se avanzó con una tesis de maestría en TIAE, en cooperación con una investigadora de la Universidad de Zaragoza que se aborda el uso de videojuegos como método de evaluación sigilosa. Se presentó una revisión sistemática sobre los videojuegos como método de evaluación sigilosa (Ferreyra, Sanz, & Coma-Roselló, 2024).

En cuanto a los proyectos vinculados con la temática y los acuerdos de cooperación, el III- LIDI participa en los siguientes:

- Se cuenta con un acuerdo de colaboración con la Universidad de Zaragoza y se trabaja en forma conjunta en estas temáticas.
- Se tiene un acuerdo de cooperación con la Universidad de Islas Baleares, en particular se trabaja con el Departamento de Ciencias Matemáticas e Informática.
- Se coopera con la Universidad Nacional de Santiago del Estero con asesoría en vinculación algunos de los ejes presentados aquí.
- Se viene participando en proyectos de Innovación Docente de la Universidad de Zaragoza.
- Se ha participado del proyecto Erasmus Cap4city.
- Se participa del Programa UNITA-Geminae[1].
- Se participa de la iniciativa CIN-UNIUEAR. Alianza Universitaria Argentina Europea para la Transformación Digital
- Se participa del Proyecto "Construction and validation of a methodology for digital transformation of localities in developing countries to promote their sustainable development", aceptado en el marco de la convocatoria de proyectos de “Perez Guerrero Trust Fund (PGTF) 2023” para países miembro del grupo de 77. Participan de este proyecto Ecuador, Argentina y Cuba.

Se llevan adelantes acciones de vinculación con el medio y transferencia. En 2024 se han desarrollado numerosos talleres con escuelas donde se utilizan los juegos, entornos, y aplicaciones desarrolladas en el marco del proyecto.

4. FORMACION DE RRHH

Se aborda la formación de recursos humanos a través de la dirección de becas, tesis de doctorado, maestría y trabajos finales de especialización y de grado.
En 2024, se cuenta como resultado con: 1 tesis doctoral finalizada y 1 entregada para evaluación, 2 por finalizarse, 1 tesis de maestría finalizada, y 1 práctica profesional supervisada finalizada. Además, se entregó un trabajo de especialización y se espera su evaluación.

5. BIBLIOGRAFÍA

- Anderson-Coto, M.J., Salazar, J., López, J.L.S., Sedas, R.M., Campos, F., Bustamante, A.S., Ahn, J. (2024). Towards culturally sustaining design: Centering community’s voices for learning through participatory design. International Journal of Child-Computer Interaction 39, 100621.https://doi.org/10.1016/j.ijcci.2023.100621
- Derouech, O., Hrimech, H., Lachgar, M., & Hanine, M. (2024). A literature review of collaborative

[1] https://univ-unita.eu/Sites/UNITA/en/

virtual environments: impacts, design principles, and challenges. Journal of Information Technology Education: Research, 23, Article 11. https://doi.org/10.28945/5283

- del Gener, A, Sanz, C., & Iglesias L. (2023). Exergames: Proposal for a Gamepad to Sense Player Movements. Journal of Computer Science & Technology (JCS&T), 23(2).
- Dieser, M. P., Sanz, C. V., Zangara, A. (2023a), Hacia una comprensión dinámica y contextual de la autorregulación del aprendizaje: Propuesta metodológica para su evaluación en foros académicos. XVIII Congreso de Tecnología en Educación & Educación en Tecnología-TE&ET.
- Dieser, M. P., Sanz, C. V., Zangara, A. (2023b), Autorregulación del aprendizaje, estrategias de participación y rendimiento académico. Un estudio de su relación en comunidades virtuales de indagación. Tesis de Maestría en TIAE. Disponible en: https://sedici.unlp.edu.ar/handle/10915/158434
- Fachal, A. S., Sanz, C. & Abásolo, M.J. (2024). Posibilidades pedagógicas de los entornos virtuales 3D en el acompañamiento del aprendizaje de personas con discapacidad auditiva Tesis Doctoral, UNLP. https://doi.org/10.35537/10915/174350
- Fachal, A.S., Abásolo, M.J., Sanz, C.V. (2024), "Experiencia Educativa con Personas con Discapacidad Auditiva mediante el uso de Entornos Virtuales 3D en OpenSim,". Revista Tecnología en Educación & Educación en Tecnología-TE&ET, ISSN 1850-9959.
- Ferreyra, P., Coma-Roselló, T., & Sanz, C. V. (2024). Revisión sistemática sobre videojuegos como método de evaluación sigilosa para analizar rasgos de personalidad. In XIX Congreso Nacional TE&ET 2024
- García, P., Artola, V., & Sanz, C (2024). "Herramientas de Inteligencia Artificial en procesos educativo". Propuesta de tesis de Maestría en TIAE aprobada por la Fac. Informática, UNLP.
- Gros, B., & Durall, E. (2020). Retos y oportunidades del diseño participativo en tecnología educativa. Edutec. Revista Electrónica De Tecnología Educativa, (74), 12-24. https://doi.org/10.21556/edutec.2020.74.1761
- Gubaro, M., Sanz, C., & Artola, V. (2024) Juegos Serios Educativos Digitales Orientados a la Enseñanza de la Matemática en Educación Secundaria. Estado del Arte. CACIC 2024. Pág.379-389
- Honcharuk, M. & Gonzalez, A. (2024). Estrategias de aprendizaje con inmersión tecnológica en contextos rurales (Doctoral dissertation, Universidad Nacional de La Plata).
- Iglesias, L. (2024). Informe técnico sobre el pipeline de trabajo de IDEAS: "Flujos de trabajo de integración y entrega continua".
- Lombardo, S, & Sanz, C. (2024). "Aportes al desarrollo y actualización del entorno virtual de enseñanza y aprendizaje Ideas. Integración de un chatbot para estudiantes y docentes". Informe final de Práctica profesional supervisada de la carrera Ing. en Computación. Fac. Informática. UNLP
- Rikkerink, M., Verbeeten, H., Simons, R. J., & Ritzen, H. (2016). A new model of educational innovation: Exploring the nexus of organizational learning, distributed leadership, and digital technologies. Journal of Educational Change, 17(2), 223-249
- Sanz, C., Artola, V. & Ibañez, B. (2024) EmpoderAR un Juego Serio de Realidad Aumentada para Acercar a Jóvenes a la Ciencia, Interacción Revista digital de AIPO, 5(1), 6-13
- Sanz, C; Gonzalez, A.; Zangara, A.; Ibañez E.; Iglesias L.; De Giusti Armando. A Virtual Learning Enviroment Development. Some experiences at the University of La Plata. Journal of Computing and Information Technology. Zagreb: Universidad de Zagreb. 2006 vol.14 n°2. p11 - 14. Issn 1330-1136.
- Sanz, C., Coma-Roselló, T. Aguelo, A., Álvarez, P y Baldassarri, S. (2023) "Model and Methodology for Developing Empathy: An Experience in Computer Science Engineering," in *IEEE Trans. on Education*, doi: 10.1109/TE.2022.3231559.
- van Leeuwen, A. & Rummel, N. (2020) Comparing teachers' use of Mirroring and advising dashboards. In: Proc. of the Tenth International Conf. on Learning Analytics & Knowledge, Frankfurt Germany, pp. 26–34. https://doi.org/10.1145/3375462.3375471
- Vásquez-Bermúdez, M., Sanz, C., Zangara, M. A., & Hidalgo, J. (2024). Evaluation of the Tool Visualization of Interaction Tracking in Collaborative Work in a Forum Activity in Moodle. In International Conference on Technologies and Innovation (pp. 200-210). Cham: Springer Nature.
- Vera-Mora, G., Sanz, C. V., Coma-Roselló, T., & Baldassarri, S. (2024). Model for Designing Gamified Experiences Mediated by a Virtual Teaching and Learning Environment. Education Sciences, 14(8), 907. https://doi.org/10.3390/educsci14080907
- Violini, M. L., Sanz, C. V. & Pesado, P. M. (2023). Proceso de diseño y desarrollo del framework MarCOA para la creación de Objetos de Aprendizaje. Argentina. Hurlingham, Buenos Aires. 2023. Libro. Artículo Completo. Congreso. XVIII Congreso Tecnología en Educación & Educación en Tecnología. UNAHUR.

LABORATORIO REMOTO

Alejandro Martín Rapallini amrapallini@udc.edu.ar (1), José Antonio Rapallini jarapallini@udc.edu.ar (1) (2), Leticia Baez lmbaez@udc.edu.ar (1), Romina Iralde criralde@udc.edu.ar (1)

(1)Universidad de Chubut, (2)Centro CODAPLI UTN–CIC

RESUMEN

Este trabajo presenta el desarrollo de laboratorios remotos en la Universidad del Chubut (UDC), orientado a mejorar la accesibilidad a experiencias prácticas en carreras tecnológicas y del ámbito de la salud, con especial foco en la Licenciatura en Enfermería. En un contexto donde la universidad cuenta con extensiones áulicas distribuidas en una amplia geografía, la implementación de laboratorios remotos permite superar barreras de infraestructura y garantizar equidad en el acceso a tecnologías para la formación. Además, se destaca el impacto social de este proyecto, al fortalecer la formación práctica de estudiantes en carreras con fuerte impacto comunitario.

Palabras clave: laboratorios remotos, tecnología educativa, acceso equitativo, enfermería, redes.

CONTEXTO

En el marco del Programa creado en marzo de 2024, en la universidad de Chubut sobre Nuevas Tecnologías (Resolución N° UDC[1]), donde se promueve, la investigación y la vinculación tecnológica con el territorio, desarrollando líneas de extensión con la comunidad a través del protagonismo de docentes, estudiantes y graduados en principio de las carreras de Licenciatura en Redes y Telecomunicaciones (LRyT), Tecnicatura Universitaria en Desarrollo de Software (TUDS), Tecnicatura Universitaria en Energías Renovables (TUER), que tienen como objetivos específicos:

1. Promover líneas de investigación a través de programas y proyectos de investigación, en articulación con otras instituciones.
2. Articular las actividades de investigación con los requerimientos del medio productivo y académico.
3. Promover la Formación de Recursos Humanos especializados en los objetivos del Programa, fomentando la participación de Docentes en investigación y transferencia.
4. Transferir los resultados al medio productivo y académico, realizando publicaciones científicas y técnicas, participando en congresos, jornadas, seminarios, etc., incentivando el desarrollo de pensamiento computacional y estimulando la actividad innovadora y lógica.
5. Propiciar proyectos de extensión tecnológicos con la comunidad.
6. Contribuir a la innovación y el desarrollo de nuevas tecnologías.
7. Generar una Incubadora de Empresas UDC, que siente las bases para el desarrollo del emprendedorismo tecnológico en la región y trabaje activamente para fortalecer la comunidad emprendedora y la innovación en la universidad y en toda la región.

Pautas que permiten generar proyectos con carreras tecnológicas y del ámbito social de la misma UDC como lo es la Lic. en Enfermería

y con otras instituciones como es el caso del presente trabajo.

1. INTRODUCCIÓN

Las universidades con sedes dispersas enfrentan el desafío de brindar igualdad de oportunidades de acceso a laboratorios y herramientas de práctica. En la UDC, la creación de un sistema de laboratorios remotos surge como una solución innovadora para garantizar que estudiantes de distintas localidades [Fig 1] puedan acceder a experiencias prácticas esenciales.

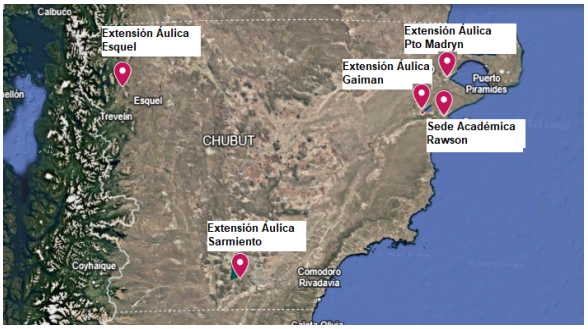


Fig 1: Localización de unidades UDC

Este proyecto [2] se diferencia de otras iniciativas similares por su aplicación en carreras de impacto social, como la Licenciatura en Enfermería, donde el acceso a simulaciones y ensayos prácticos es crucial para la formación de profesionales en contextos donde la presencia en laboratorios físicos no es viable. Además, se contempla su aplicación en la Licenciatura en Redes y Telecomunicaciones, facilitando la experimentación con equipos reales a distancia..

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

2.1. Hipótesis de trabajo:

El acceso equitativo a laboratorios y herramientas prácticas es fundamental para la calidad de la formación universitaria en carreras tecnológicas y de la salud. La implementación de laboratorios remotos permitirá reducir las barreras geográficas y presupuestarias que limitan la educación práctica, asegurando que todos los estudiantes, independientemente de su ubicación, puedan acceder a experiencias de aprendizaje de alta calidad.

Se plantea que el uso de tecnologías remotas para la experimentación en entornos controlados mejorará la calidad del aprendizaje en la Licenciatura en Enfermería y en Redes y Telecomunicaciones, permitiendo una formación más flexible, eficiente y accesible.

2.2 Propuesta de trabajo:

La propuesta se centra en el diseño, desarrollo e implementación de una infraestructura de laboratorio remoto [3],[4],[5],[6],[7] en la UDC, con un enfoque especial en la accesibilidad y adaptabilidad para diferentes carreras.

Los ejes principales de la propuesta incluyen:

- Desarrollo de una plataforma web que permita a los estudiantes acceder y operar remotamente equipos experimentales.
- Integración de sensores, actuadores y sistemas de control remoto para la realización de experimentos en tiempo real.
- Evaluación del impacto en la experiencia de aprendizaje mediante pruebas piloto en diversas asignaturas.
- Establecimiento de vínculos con otras instituciones para la colaboración en el desarrollo y uso compartido de laboratorios remotos.
- Esta propuesta busca garantizar una formación práctica equitativa y de calidad, permitiendo que la UDC se convierta en un referente en la implementación de laboratorios remotos para la educación superior.

3. RESULTADOS OBTENIDOS/ ESPERADOS

Es de esperar, resultados del proyecto, como:

- Accesibilidad equitativa: Los laboratorios remotos permitirán a estudiantes de todas las extensiones áulicas acceder a experiencias prácticas, independientemente de su ubicación geográfica.
- Fortalecimiento de la formación en Enfermería: A través de simulaciones y ensayos remotos, los estudiantes de Enfermería podrán desarrollar habilidades esenciales sin depender exclusivamente de laboratorios físicos.
- Optimización de recursos: Al centralizar el acceso a equipamiento costoso, se reducirá la necesidad de replicar laboratorios en cada sede, minimizando costos de adquisición y mantenimiento.
- Colaboración interinstitucional: La integración de la UDC en redes de laboratorios remotos fortalecerá la cooperación con otras universidades e instituciones tecnológicas.
- Desarrollo de competencias digitales: Los estudiantes adquirirán habilidades en el uso de plataformas remotas, esenciales en el contexto de la educación y el trabajo modernos.

3.1 Primeros resultados:

Durante la fase inicial del proyecto, se han logrado avances significativos:

- Desarrollo de la plataforma piloto: Se ha implementado una versión inicial del laboratorio remoto con funcionalidades básicas de acceso y control de equipos a distancia.[8][Fig: 2]
- Pruebas con estudiantes y docentes: Se han realizado ensayos con grupos reducidos para evaluar la usabilidad y efectividad del sistema. [Fig. 3]
- Selección de equipamiento clave: Se han identificado los dispositivos tecnológicos esenciales para las prácticas en Redes y Enfermería.
- Vinculación con el entorno académico: Se han establecido contactos con otras universidades e instituciones para intercambiar experiencias y conocimientos sobre laboratorios remotos.[9]
- Evaluación preliminar del impacto: Se han recogido datos sobre la experiencia de los usuarios, destacando mejoras en la accesibilidad y la flexibilidad en el aprendizaje.

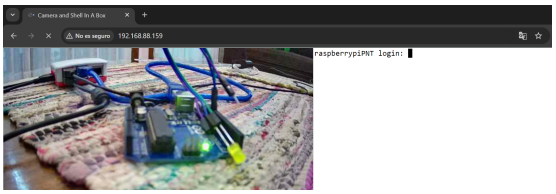


Fig 2: Desarrollo del Primer laboratorio Remoto


Fig 3: Reuniones de trabajo.

Estos primeros resultados confirman la viabilidad del proyecto y sientan las bases para su expansión y consolidación en la UDC.

4. FORMACIÓN DE RECURSOS HUMANOS

Como lo indican los objetivos del proyecto, uno de los aspectos más importantes es lograr la formación de un equipo de trabajo interinstitucional con amplios conocimientos sobre la temática, que permitirá el fortalecimiento de lazos entre escuelas de la misma Universidad y entre universidades e instituciones participantes.

5. AGRADECIMIENTOS

- Profesionales de la Universidad del Chubut [UDC], del Programa sobre Nuevas Tecnologías: Dra. M. Florencia

Torres [Coord. TUER], Ing. Nicolás Gonzalez Lloyd [Coord. LRyT], Lic. Julio Casco [Coord. TUDS], Ing. Carla Romina Iralde [Docente UDC], Ing. Matías Sosa [Docente UDC], Ing. Sebastián Azparren [Docente UDC] y Dra. Mónica Kozykariski [Directora Escuela de producción, Ambiente y desarrollo sostenible]

- Asesoramiento de profesionales del Centro de Codiseño Aplicado [CODAPLI] perteneciente a la Universidad Tecnológica Nacional, Facultad Regional La Plata y Comisión de Investigaciones Científicas de la Provincia de Buenos Aires, Argentina
- Estudiantes de la carrera Tecnicatura Universitaria en Redes y Telecomunicaciones extensión áulica Esquel, Lucas Varela, Nazareno Campos y Ariel Nehuen.

6. BIBLIOGRAFÍA

[1] Programa de Nuevas Tecnologías, Ing. Nicolás González Lloyd, Mg. Ing. Alejandro Rapallini, Lic. Julio César Casco, Dra. M. Florencia Torres, Ing. José A. Rapallini. - UDC - dic 2023 – sedes Esquel y Rawson, Chubut.

[2] Convocatoria UDC a Proyectos de Investigación con financiamiento. Laboratorio Remoto. Rapallini, Alejandro Martín; Rapallini José Antonio (UDC / CODAPLI),

[3] Á. Villalobos M., R. Romero A., “Laboratorios remotos en educación superior: Una revisión bibliográfica sistematizada,” Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología, no. 35, pp. 58-68, 2023. doi:10.24215/18509959.35.e7

[4] A. F. Almarshoud, “The advancement in using remote laboratories in electrical engineering education: a review,” Eur. J. Eng. Educ., vol. 36, no. 5, pp. 425–433, Oct. 2011, doi: https://doi.org/10.1080/03043797.2011.604125

[5] E. Guimarães, E. Cardozo, D. H. Moraes, and P. R. Coelho, “Design and implementation issues for modern remote laboratories,” IEEE Trans. Learn. Technol., vol. 4, no. 2, pp. 149–161, 2011, doi: https://doi.org/10.1109/TLT.2010.22

[6] N. Lima, G. Alves, C. Viegas, and I. Gustavsson, “Combined efforts to develop students experimental competences,” in 3rd Experiment International Conference: Online Experimentation, 2015, pp. 243–248, doi: https://doi.org/10.1109/EXPAT.2015.7463273

[7] Domínguez González, M., Fuertes Martínez, J.J., Reguera Acevedo, P., Prada Medrano, M.A., Alonso Castro, S., Morán Álvarez, A. (2009). ¨Desarrollo de un laboratorio remoto para la formación vía internet en automático¨. Premio a la Innovación en la Enseñanza 2009, 11-102

[8] Informes técnicos de Laboratorio PNT Esquel, setiembre/octubre 2024, Esquel, UDC

[9] Alejandro M. Rapallini, Marcelo Zabaljauregui, Héctor Hugo Mazzeo, Omar E. Rodríguez, José A. Rapallini. (2024) ´Desarrollo de una plataforma de laboratorios remotos con microcontroladores´.Cuarto Congreso Virtual de Microcontroladores y sus Aplicaciones "Las nuevas tendencias tecnológicas" - 30 de septiembre de 2024 al 3 de octubre 2024 - UTN Facultad Regional Paraná [AJEA ISSN:2683-8818]

OBSERVATORIO DE TECNOLOGÍAS APLICADAS A LA EDUCACIÓN (FACENA-UNNE)

María Viviana Godoy Guglielmone (mvgodoy@exa.unne.edu.ar), Beatriz Castro Chans (beatrizcc@exa.unne.edu.ar), Guillermo Andrés Arduino (gaarduino@exa.unne.edu.ar), María Lorena Guastavino Mosna (maria.lorena.guastavino@comunidad.unne.edu.ar), Marta Mabel Meza (martameza@exa.unne.edu.ar), Paola Insaurralde (paola.insaurralde@exa.unne.edu.ar), Flavia Sabrina Moreiro (flaviamoreiro@exa.unne.edu.ar), Flavia Pucheta (flaviapucheta@exa.unne.edu.ar).

Observatorio de Tecnologías Aplicadas a la Educación, Departamento de Informática, Facultad de Ciencias Exactas y Naturales y Agrimensura, Universidad Nacional del Nordeste

Eje: Tecnología Informática Aplicada en Educación

Resumen

El Observatorio de Tecnologías Aplicadas a la Educación FaCENA – UNNE tiene como propósito: producir, organizar, evaluar y procesar información en torno a la utilización de tecnologías digitales aplicadas al desarrollo de ofertas educativas de dicha Unidad Académica.

Se emplean categorías de análisis, métodos y herramientas conforme los estándares aceptados y en concordancia con los procedimientos de Evaluación y Monitoreo del Sistema de Educación a Distancia de la UNNE (SIED).

Se trata de una investigación aplicada encuadrada bajo el método cuanti-cualitativo. El objetivo de la investigación se ubica en la tipología exploratorio-descriptiva, por cuanto está orientado a identificar elementos significativos y, luego, describir con precisión el caso contextualmente situado. Los instrumentos de investigación utilizados son: encuestas, entrevistas y análisis de documentos.

El conocimiento producido opera como retroalimentación del propio sistema para la toma de decisiones al interior de la Unidad Académica en un escenario en el cual las Instituciones de Educación Superior en todo el mundo se orientan a modelos de enseñanza híbridos y flexibles, tendencia profundizada a partir de la pandemia por COVID-19.

Palabras Clave: educación superior, tecnologías educativas, juegos serios, inteligencia artificial

Contexto:

El proyecto se inserta en el Grupo de Investigación *Sistema de información y TI: modelos, métodos y herramientas* acreditado por la Secretaría General de Ciencia y Técnica de la Universidad Nacional del Nordeste (21F010 – SGCyT – UNNE).

Este proyecto surge en el marco de la creación y puesta en funcionamiento del Área de Educación Virtual de la Facultad de Ciencias Exactas y Naturales y Agrimensura de la Universidad Nacional del Nordeste, aprobada por el Consejo Directivo de la misma y cuyas funciones se detallan en la Resol. N° 356/19.

1. Introducción:

Se trata de un proyecto de investigación aplicada que tiene como propósito la construcción y aplicación de modelos de evaluación en torno al uso y aplicación de tecnologías educativas cuya aplicación provea información válida y sistemática sobre el campo.

Además, busca promover el desarrollo de tecnologías que surjan de dichas áreas a través de la incorporación de recursos humanos que se encuentren realizando trabajos finales de graduación, becas y/o pasantías.

El Observatorio articula estrechamente con el Área de Educación Virtual de la Facultad de Ciencias Exactas y Naturales y Agrimensura –

UNNE (FaCENA). En este sentido, el Observatorio aporta a la permanente actualización, incorporación de cambios e innovación vinculados con actividades de docencia, investigación, desarrollo experimental y transferencia con miras a fomentar el desarrollo del conocimiento y la formación de recursos humanos considerando especialmente las particulares condiciones de accesibilidad en la región NEA.

El proyecto tiene como objetivos específicos:

• Recolectar, organizar, evaluar y procesar información sobre las tecnologías digitales utilizadas en los trayectos formativos de la FaCENA-UNNE.

• Proveer información significativa y sistematizada a las autoridades y equipos de gestión de FaCENA para la toma de decisiones en materia de escenarios futuros de educación virtual.

• Proporcionar información válida y sistematizada a la comunidad docente e investigadora de FaCENA, al Sistema Institucional de Educación a Distancia de la UNNE, a entidades gubernamentales de nivel provincial y nacional, redes interinstitucionales y otras instituciones interesadas en la temática.

• Identificar demandas formativas del campo científico – tecnológico en la Región que puedan resolverse con mediación de tecnologías, como estrategia para la inclusión y la equidad en el acceso a la educación superior, continua y de calidad.

Las Tecnologías de la Información y de la Comunicación aplicadas a los procesos de enseñanza y de aprendizaje constituyen una temática de gran relevancia en el ámbito de la educación superior, debido a que las mismas modifican y aportan significativamente a las funciones sustantivas de las universidades.

En esta línea, Igarza (2021, p. 28) partiendo de lo acontecido en la educación a partir de la pandemia por COVID19, afirma que el futuro de la Educación Superior "se vincula con dos estrategias: por un lado, la gestión y la gobernanza institucional que asegure una transición administrativa y simbólica lo menos traumática posible. Por otra parte, las estrategias de apoyo, un andamiaje que debe asegurar el desarrollo de políticas, procedimientos y acciones en tres planos: i) el técnico-pedagógico ii) el pedagógico y metodológico, y iii) el temático disciplinario, con personas formadas para tales finalidades, y una política de desarrollo tecnológico sustentable, modular y escalable que comprenda lineamientos sobre la interoperabilidad técnica y sobre el programa de inversión y obsolescencia." En otras palabras, la reorientación de las prácticas, muchas de ellas basadas en el uso de tecnologías debe responder a políticas institucionales diseñadas en base a información.

En la misma línea, Lion, Kap y Ferrareli (2023) afirman que *"El desarrollo de tecnologías y nuevas formas de comunicación han dado lugar a un momento de alfabetismos fluidos e hibridaciones, lo que requiere que las universidades construyan nuevas estrategias de enseñanza para preparar a los sujetos en la articulación de nuevos y viejos aprendizajes. A esto se suma la imposibilidad de abordar las prácticas digitales y analógicas por separado, una mixtura de haceres y saberes en donde convergen interrogantes sobre la manera de articular la enseñanza en la universidad con aprendizajes formales e informales."*

Según lo expresa el Banco Interamericano de Desarrollo (2009) un Observatorio es un proceso de acompañamiento de largo plazo para la reflexión y sistematización de aprendizajes que se realiza durante la implementación de una experiencia o proyecto e involucra a todos sus actores, especialmente miembros del equipo de proyecto, unidades ejecutoras y beneficiarios. La idea de observatorio virtual es un concepto más innovador, basado en la colaboración en línea y el uso de fuentes de información de diversa índole, que se alimenta habitualmente con información de una comunidad afín; después de su procesamiento y análisis, los resultados se ponen a disposición de la propia comunidad. (Castañeda, 2006).

Un observatorio puede asumir enfoques diferentes: el primero, relacionado con almacenes de información y generación de informes, y el segundo, con formas más

dinámicas sustentadas en la colaboración que estimulan la comunicación y promueven la reflexión.

La gestión de la información en las universidades nacionales constituye una valiosa herramienta para la planificación y toma de decisiones. En este sentido, disponer de información de calidad que represente la realidad institucional constituye la mejor inversión y representa una fuente de poder de enorme importancia. El observatorio que se propone en este PI se enmarca institucionalmente en los requerimientos de las políticas educativas en nuestro país, a raíz de la Resolución Nº 2641/17-E (reemplazada por la Res. 25999/23 ME) que exige a las instituciones universitarias que quieran implementar ofertas a distancia deben, entre otras cosas, validar su Sistema Institucional de Educación a Distancia y prever dispositivos para su autoevaluación y generación de conocimientos relativos al campo.

Lion (2020, p.98), haciendo referencia a los retos para la formación docente en escenarios actuales, manifiesta que: *"las investigaciones que han sido prolíferas en este entrecruzamiento de tecnologías y formación, articuladas con [el reto anterior de] las políticas y una participación de los sujetos en formación, especialistas, decisores, actores diversos del sistema; son relevantes para una toma de decisiones consensuada para gestar cambios en nuestras instituciones educativas"*.

Previamente a la expansión de las TIC, ya se empleaban otros recursos tecnológicos aplicados a los procesos educativos. En la actualidad el alcance y las consecuencias de la convergencia digital, el uso de modelos de simulación y aprendizajes no lineales basados en el hipertexto, sumado a los requerimientos de formación en competencias digitales (informáticas e informacionales) expresados por el mercado laboral han modificado los paradigmas y las prácticas educativas.

Lion (2020, p.26) señala que: *"las tecnologías sostienen nuevas formas de producción y circulación del conocimiento en las que se valora la polifonía de voces, el trabajo en colaboración y las propuestas revisadas de autoría y de construcción colectiva de redes sociales."*

Las tecnologías digitales promueven practicas tecnoculturales en las instituciones educativas. Poseen particularidades que las diferencian de otros recursos (Coicaud, 2019). En los últimos años se distinguen las TIC, de las TAC y las TEP.

Las TIC no solo han reconfigurado los roles de quienes enseñan y de quienes aprenden, sino que los espacios educativos han pasado a ser intervenidos por las herramientas tecnológicas y han llegado, en algunos casos, a nuevos conceptos de aula, y en otros, a espacios intangibles, concebidos por lo digital. Las TAC, tecnologías para el aprendizaje y el conocimiento, es un concepto que sirve para identificar las tecnologías impulsadas al fortalecimiento del proceso de enseñanza-aprendizaje. Con ello, se entiende cómo, a partir de estas mediaciones tecnológicas, la escuela promueve en los estudiantes una postura de crítica y análisis, constructiva y responsable, difundidas o socializadas mediante las TEP, a saber, las tecnologías de empoderamiento y participación, por ser estas, en última instancia, el final de un proceso educativo que se proyecta del aula al entorno social y que logra la construcción de un conocimiento colectivo de alto impacto. (Latorre Iglesias, Castro Molina y Potes Comas. (2018, p. 37-38)

Para Cabero, Barroso y Llorente (2015) las tecnologías educativas pueden ser comprendidas y aplicadas de diferentes formas. Pueden ser observadas desde una perspectiva "micro", atendiendo a los medios y recursos tecnológicos, como desde un punto de vista más amplio, que es el referido a la planificación de la instrucción.

Las perspectivas que las tecnologías de la información y la comunicación presentan para las próximas décadas implican un avance importante hacia los dispositivos móviles. Las tecnologías móviles atribuirán un carácter más nómade a las escuelas, promoviendo en mayor medida la comunicación fluida entre educadores y estudiantes de diferentes lugares del planeta. (Coicaud, 2019. p, 23).

Por otra parte, es necesario recuperar los aportes de los estudios sobre apropiación social de las TIC. Estos se remontan a los años 90, en los cuales se ha consolidado una línea de investigación, en donde se articulan desarrollos teóricos de diferentes disciplinas (Gonzalo, 2014).

El análisis de las características y dimensiones de la variable 'uso', se relaciona con los estudios de 'apropiación social de la tecnología' o 'apropiación tecnológica'. En términos generales, el objetivo de estos estudios es investigar acerca de las formas y los contextos en los que diferentes grupos sociales, además de conocer y tener las tecnologías, 'se apropian' -o no- de ellas y las utilizan de una manera efectiva en sus actividades cotidianas (Echeverría: 2008; Toboso-Martín: 2014)

Pardo Kuklinski y Cobo (2020) expresan la necesidad de que las universidades, especialmente las públicas, estén al mando de la producción de conocimiento en términos de "inteligencia organizacional" para la toma de decisiones y liderazgo de los procesos de cambio institucional; afirman que: *"El conocimiento científico de excelencia, vanguardista y en conexión con el saber producido en diferentes centros de investigación del mundo, serán la mejor forma de alcanzar*
soluciones a los retos actuales."

El presente proyecto tiene, además de los antecedentes en términos de conocimiento sobre el observatorio como dispositivo de producción de conocimiento, las tecnologías aplicadas a la educación y las nociones ligadas a la apropiación de tecnologías, un contexto normativo - institucional que funciona como marco de referencia, conformado por: la LES, la RM 2641/17 ME que regula la Educación a Distancia en Argentina, el propio SIED-UNNE y "Procedimientos para el monitoreo y evaluación del Sistema Institucional de Educación a Distancia de la UNNE"
(Resoluciones Nº 221/18 CS-UNNE y Nº 915/18 CS-UNNE).

2. Líneas de Investigación y Desarrollo

Actualmente el proyecto de investigación está enfocado en el análisis de la apropiación y uso de inteligencia artificial generativa, las necesidades de formación de los usuarios en materia de IA y los desafíos que presenta esta tecnología en el campo de la educación.

Además, a partir de la expansión del uso de herramientas y estrategias de gamificación en el aula, ha desarrollado algunas experiencias de juegos serios, especialmente orientados a estudiantes de los primeros años para evaluar su percepción y resultados.

Por otra parte, se analizan la accesibilidad a tecnologías para el desempeño académico y los requerimientos de formación en tecnologías aplicadas a la educación por parte de docentes y estudiantes de nivel universitario.

Finamente, se busca contribuir a la divulgación de tecnologías educativas innovadoras y buenas prácticas en el uso de tecnologías educativas a través de seminarios, cursos de capacitación y posgrado y actividades de divulgación (FACIENCIA, Jornadas, Congresos).

3. Resultados obtenidos / esperados

El proyecto proporciona a la Facultad, en primera instancia, información sistematizada sobre la inclusión efectiva de las TIC en las prácticas de enseñanza-aprendizaje, metodologías aplicadas y resultados obtenidos a través de documentos, informes y publicaciones.

Estos materiales constituyen un recurso para la producción de recomendaciones y la toma de decisiones en diferentes niveles (docencia y gestión). De este modo, contribuye a la línea de investigación en Educación mediada por tecnologías aprobada por el Consejo Superior, Resolución N°2008/21 y el fortalecimiento del SIED-UNNE (Resolución N°221/018).

También aporta a la identificación de áreas de vacancia en términos de desarrollo de conocimiento y tecnologías (i+D) orientadas respecto de las tecnologías aplicadas a la educación y analíticas de aprendizaje a través de la realización de encuestas periódicas a docentes y estudiantes.

En otro orden, desde el proyecto, se elaboran materiales educativos en diferentes formatos

(infografías, videos, etc.), la organización de seminarios y jornadas para la conformación y fortalecimiento de la comunidad de práctica de docentes, por ejemplo: las Primeras Jornadas de Educación y TIC de FaCENA – UNNE. «El desafío de la enseñanza virtual en carreras científico-tecnológicas. Entre tradiciones y nuevas realidades» (2021), webinarios sobre el uso de tecnologías (2020 – 2024), el Ateneo de Inteligencia Artificial de la FaCENA – UNNE (2023) y la participación en la Diplomatura en Inteligencia Artificial aplicada la Educación (Nivel Básico) en 2024.

4. Formación de Recursos Humanos

El equipo de investigación se encuentra formado por: una Magister en Informática, una Magister en Ciencias Sociales y Humanas en elaboración de tesis doctoral sobre Sistemas Institucionales de Educación a Distancia, un Doctorando en Informática (en elaboración de tesis), una Especialista en Educación Superior cursando la Maestría en Informática, una Profesora de Matemática y una Lic. En Relaciones Laborales realizando su trabajo final de Especialización en Educación con Tecnologías, una Especialista en Educación próxima a defender) su tesis de Maestría en Educación. (2025) y una Licenciada en Sistemas. Además, se han dirigido dos tesis de grado ya defendidas y en proceso de ejecución de otra tesis de grado en tecnologías educativa.

5. Bibliografía:

Angulo Marcial, N. (2009) *¿Qué son los observatorios y cuáles son sus funciones?* En: Innovación Educativa, vol. 9,núm. 47, abril-junio, 2009, pp. 5-17. México: Instituto Politécnico Nacional.

Banco Interamericano de Desarrollo (2009) *Guía metodológica para el diseño e implementación de Observatorios de Experiencias Innovadoras.*

Cabero, J. Barroso, J. & Llorente, M.C. (2015). E-actividades para la formación del profesorado en tecnologías de la información y comunicación en el Proyecto Dipro 2.0 Sevilla: Grupo de Investigación Didáctica de la Universidad de Sevilla

Castañeda de León, L.M. (2006). Observatorio virtual: más que un portal. En: Entér@te en–línea. UNAM. Dirección General de Servicios de Cómputo Académico. Año 5, Núm. 46, febrero de 2006. http://www.enterate.unam.mx/Articulos/2006/febrero/observa.htm

Cocaiud, S. (2019) *Potencialidades didácticas de la inteligencia artificial. Videojuegos, realidad extendida, robótica y plataformas. Mediaciones tecnológicas para una enseñanza disruptiva*. Buenos Aires: NOVEDUC.

Echeverría, J. (2008). Apropiación social de las tecnologías de información y la comunicación. En: Revista Iberoamericana de Ciencia, Tecnología y Sociedad, n.º 10, vol. 4, Buenos Aires.

Gonzalo, D. (2014) Una aproximación conceptual a la "Apropiación Social" de TIC. En: Revista Question, Vol. 1, N.° 43(julio-septiembre de 2014), UNLP.

Igarza, Roberto (2021) *Presencias imperfectas. El futuro virtual de lo social*. Buenos Aires: La Marca Editora.

Latorre Iglesias, E., Castro Molina, K. y Potes Comas. I. (2018) *Las TIC, las TAC y las TEP: innovación educativa en la era conceptual*. Bogotá: Universidad Sergio Arboleda.

Pardo Kuklinski, H. y Cobo, C. (2020). Expandir la universidad más allá de la enseñanza remota de emergencia Ideas hacia un modelo híbrido post-pandemia. Barcelona:Outliers School.Latorre

Iglesias, E., Castro Molina, K. y Potes Comas. I. (2018) Las TIC, las TAC y las TEP: innovación educativa en la era conceptual. Bogotá: Universidad Sergio Arboleda.

Lion, C., Kap, M. y Ferrareli, M. (2023) *Universidades desafiadas: alfabetismos fluidos, hibridaciones y nuevas estrategias de enseñanza*. Revista Educación Superior y Sociedad, 2023, VOL.35 N.º 2, 130-155. https://doi.org/10.54674/ess.v35i2.768

UNNE (2018) *Sistemas Institucional de Educación a Distancia de la Universidad Nacional del Nordeste*. Resolución N° 221/18 Consejo Superior.

Procesos de apropiación y uso de tecnologías móviles en Río Negro. Primeros pasos

Edith Lovos[1]; Lucía Cantamutto[4], Alejandra Marin Aranda[1], Martín Goin[2], Pablo Fabian Carranza[3], Jose Ignacio Muñoz[1], Mariana Roche[4], Emanuel Llancapan [1]

[1] Universidad Nacional de Río Negro, Sede Atlántica, Viedma, Río Negro
[2] Universidad Nacional de Río Negro, Sede Andina,Bariloche, Río Negro
[3] Universidad Nacional de Río Negro, Sede Alto Valle y Valle Medio, Allen, Río Negro
[4]CONICET, Universidad Nacional de Río Negro, CURZA -Universidad Nacional del Comahue
Centro Interdisciplinario en Derechos, Inclusión y Sociedad (CIEDIS)
{elovos, lcantamutto,mmarin,mgoin,pcarranza,jmunoz}@unrn.edu.ar

RESUMEN

En este trabajo se presenta los lineamientos generales de un proyecto de investigación aprobado y financiado por la Universidad Nacional de Río Negro (UNRN) que propone indagar sobre los procesos de apropiación de tecnologías digitales móviles, por parte de los estudiantes de los diferentes niveles educativos y modalidades que tienen lugar en la provincia de Río Negro, y que los mismos adoptan en su quehacer cotidiano. Contar con esta información, permitirá avanzar en la configuración de un Observatorio de tendencias en el uso de tecnologías móviles a nivel regional, que sea un aporte para la toma de decisiones por parte de las instituciones educativas así como de los actores políticos, sociales y de la comunidad en general.

Palabras clave: TECNOLOGÍAS MOVILES, PROCESOS DE APROPIACIÓN, INSTITUCIONES EDUCATIVAS, RÍO NEGRO .

CONTEXTO

Este trabajo presenta un resumen de los primeros meses de ejecución del proyecto de investigación denominado “Procesos de apropiación y uso de tecnologías móviles en Río Negro (Argentina): Desafíos para las instituciones educativas” (PI-UNRN40C1149), aprobado y financiado por la UNRN . El mismo tiene una duración bianual y comenzó a finales de julio de 2024. El proyecto forma parte de las actividades de investigación que se realizan en el Centro Interdisciplinario de Estudios sobre Derechos Inclusión y Sociedad (CIEDIS) perteneciente a la Sede Atlántica de UNRN, específicamente dentro de la línea de investigación sobre epistemología, investigación e innovación docente.

1. INTRODUCCIÓN

A partir de los avances tecnológicos y las posibilidades de acceso a los mismos, su uso e integración es cada vez más frecuente e inclusive naturalizada en diferentes ámbitos de la vida cotidiana de las personas en general. En el caso de los espacios educativos y culturales la penetración de dispositivos móviles, permite el acercamiento a las denominadas tecnologías disruptivas, entre ellas la realidad aumentada (RA), la realidad virtual (RV), y la inteligencia artificial (IA) entre otras, y abre posibilidades a nuevas formas de interacción con los

contenidos generados y también entre los usuarios. En el caso particular de la IA, su penetración retoma debates en relación a los límites y regulaciones que demanda su uso en los diferentes ámbitos de la vida cotidiana y con especial atención en los hábitos de uso por parte de niños y adolescentes. Así es posible encontrar informes como el de Charisi et al.,(2022), donde entre los riesgos que suponen para las infancias la IA resaltan: los sistemas de recomendaciones, los agentes de conversación y/o los robots sociales, en cuanto a violación de la privacidad (recuperación y procesamiento de datos), la presencia de estereotipos y/o la dicción no adecuada al grupo etario, sumado a la falta de garantías que posibiliten a futuro el "derecho al olvido". Sumado a ello, Corica (2023), pone en debate la "capacidad de decisión" de los individuos en un contexto donde la IA se integra de forma transparente y sin advertencia, y plantea "¿Es una decisión influenciada o construida por un algoritmo de inteligencia artificial?". En línea con este planteo, para Urresti et al.;(2015: 20) "a diferencia de las generaciones de jóvenes de otras épocas, los jóvenes actuales se integran en el mundo de la comunicación digital sin filtros previos ni membranas protectoras, haciendo del contexto social crecientemente digitalizado el centro de su experiencia de vida". Por otra parte, Silvia Coicaud (2019) plantea la necesidad de aprovechar el potencial didáctico de estas tecnologías, revalorizando las prácticas de enseñanza y aprendizaje al estilo laboratorios de innovación. En este sentido, Ferreira y Leliwa (2023), dan cuenta de la importancia que tienen los espacios de educación tecnológica para las infancias y adolescencias, entendiendolos como el lugar, donde pueden entrar en contacto con la cultura digital desde una perspectiva reflexiva, que los invite a ser críticos individual y socialmente con la tecnología que los atraviesa en forma cotidiana.

Atendiendo a estos enfoques, se buscará a lo largo del proyecto generar conocimiento sobre los procesos de apropiación de tecnologías digitales móviles, por parte de los estudiantes de los diferentes niveles educativos y modalidades que tienen lugar en la provincia de Río Negro, y que los mismos adoptan en su quehacer cotidiano.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Desde el proyecto se propone avanzar en:

- Diseñar propuestas pedagógico - didácticas situadas en el contexto de las infancias y adolescencias de la provincia de Río Negro, que permitan desarrollar conocimientos y habilidades para comprender e intervenir las tecnologías digitales en general y las disruptivas en particular
- Diseñar y producir materiales educativos digitales, entre ellos juegos serios qué incluyan tecnologías disruptivas y permitan promover hábitos de uso seguro, prevenir y/o detectar situaciones que ponen en riesgo y/o atentan contra el bienestar de las infancias y adolescencias
- Revisar los diseños curriculares vinculados tecnologías y educación, en especial en la formación de formadores
- Realizar actividades de transferencia situadas en las instituciones públicas y privadas de nivel medio y primario de la provincia de Río Negro

Desde lo metodológico se ha optado para el desarrollo de proyecto, por un estudio de tipo exploratorio-descriptivo, siguiendo el enfoque mixto predominantemente cualitativo sin descartar aportes cuantitativos. Para definir las unidades de análisis se han tomado los aportes de Giniger y Figari (2015).

3. RESULTADOS OBTENIDOS/ESPERADOS

A continuación se presentan los resultados obtenidos hasta el momento :

- Análisis de una experiencia educativa de acercamiento y exploración de IA generativa en prácticas de lectura y comprensión de textos. La misma estuvo destinada a estudiantes universitarios de formación docente (nivel primario) de un curso virtual sobre tecnologías y educación. La experiencia permitió explorar los desafíos que provoca este tipo de tecnología, las habilidades técnicas y blandas que demanda su uso, así como también de los posicionamientos que los estudiantes (en ejercicio de la práctica docente) ponen en juego en el proceso de interacción con la IA. Un resumen de la experiencia fue aceptado para su presentación y discusión en la escuela de verano 2025 de las III Latin American STEAM Education Research Conference (UTN - Avellaneda Febrero 2025)
- Desarrollo de una conferencia en modalidad virtual sobre La intersección entre la IA (datos, entrenamiento, predicción) y el sistema educativo (aula, gestión y organismos. La misma se llevó adelante en septiembre de 2024 en la UNRN.
- En diciembre de 2024, se llevó adelante una experiencia de enseñanza y aprendizaje de conceptos básicos de programación, usando un juego unplugged. La experiencia estuvo destinada a niños y niñas de 6to grado de educación primaria de una escuela primaria de gestión privada del Sur de la provincia de Buenos Aires. En los próximos meses, se espera avanzar en un estudio comparativo con datos provenientes de una experiencia de iguales características llevada adelante en San Carlos de Bariloche.
- Se presentó una postulación a la convocatoria CIN 2024, con una propuesta qué busca indagar sobre dispositivos móviles e IA llevando adelante un análisis del repertorio de aplicaciones que utilizan los ingresantes universitarios de Río Negro.

Por último, se presentó una propuesta de extensión a la convocatoria UNRN 2024. La misma tiene como objetivo la alfabetización en datos sobre diabetes en la población de la comarca Viedma - Patagones en particular en aquellos contextos vinculados al sector educativo (escuelas, institutos, universidades, entre otros).

4. FORMACIÓN DE RECURSOS HUMANOS

Los investigadores que conforman el equipo de este proyecto, son docentes en actividad en asignaturas del área de matemáticas, literatura, programación de computadoras y tecnología educativa a través de las carreras vinculadas a ingenierías, sistemas, administración y educación en las diferentes sedes y modalidades ofrecidas por las instituciones educativas del contexto de aplicación (UNRN, UNCOMA e Institutos de formación docente de la provincia de Río Negro).

Para cerrar, es importante mencionar la importancia de la conformación interdisciplinar e interinstitucional del equipo de trabajo, como un aporte a la construcción e interpretación de los fenomenos socioeducativos sobre los que se enfoca la investigación.

5. REFERENCIAS

Charisi, V., Chaudron, S., Di Gioia, R., Vuorikari, R., Escobar Planas, M., Sanchez, M. J. I., & Gomez

Coicaud, Silvia. (2019). Potencialidades didácticas de la inteligencia artificial:

mediaciones tecnológicas para una enseñanza disruptiva (1.ª ed. papel). Noveduc, 128 pp

Corica, Luis. (27/09/2023). La era de la IA pone en juego la libertad de decisión del ser humano. Listin Diario. https://listindiario.com/la-republica/20230927/ia-pone-juego-libertad-decision-humano_774873.html

Ferreyra, Y., & Leliwa, S. (2023). Potencialidades en el aprendizaje de la Educación Tecnológica en las infancias y adolescencias. Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología, (34), 59-65.

Giniger, N y Figari, C. (2015) Taller de Metodología de la Investigación. En el marco del Doctorado de Educación de la Universidad Nacional del Comahue. Documento de cátedra.

Urresti, M., Linne, J., & Basile, D. (2015). Conexión total. Ediciones del Aula Taller.

Sobre el diseño del arte y mecánicas de un videojuego para la prevención del Abuso Sexual Infantil

Roma, Cecilia; Manghi Y Scheck, Milena; Baya, Mateo

Videojuegos para la prevención del ASI / CAETI - Centro de Altos Estudios en Tecnología Informática / Tecnología Informática / Universidad Abierta Interamericana
Av. Montes de Oca 745
(C1270AAH) - Ciudad Autónoma de Buenos Aires
Tel: tel: 54 11 4301-5323
mariacecilia.roma@uai.edu.ar

Resumen

La prevención del Abuso Sexual Infantil (ASI) requiere concientizar a todas las personas, desde niños hasta adultos. Para promover la prevención se torna fundamental inculcar hábitos de cuidado del cuerpo y de identidad a niños y niñas. En el caso de los adultos, el énfasis está centrado en brindar información para que puedan identificar posibles situaciones de riesgo y acciones correspondientes ante casos de abusos o sospecha de abuso.

Según un informe relevado por UNICEF [1] el ASI está mediado por una relación asimétrica de poder, en la cual un niño, niña o adolescente es víctima y una persona adulta o por lo menos mayor al niño, es agresora. Refiere al derecho a decidir sobre su cuerpo y sexualidad, a que sea respetada su privacidad e intimidad y además vivir sin violencias.

Con base en estas problemáticas se diseñó un proyecto para el desarrollo de un videojuego visando la prevención del Abuso Sexual Infantil.

El uso de los videojuegos como recurso informativo y educativo ha quedado demostrado por varias investigaciones por ejemplo [2], [3] y [4] entre otros.

En el caso del presente videojuego, se ha empleado un diseño específico de mecánicas y de arte con objetivo de informar de manera preferentemente gráfica y lúdica riesgos de la vida real.

Objetivo: Diseño de un videojuego cuya narrativa, mecánicas y arte se complementen para enseñar al niño acciones preventivas ante posibles situaciones de Abuso Sexual Infantil.

Palabras clave: Abuso Sexual Infantil. Videojuego. Prevención. Aprender haciendo. Niñas, Niños y Adolescentes. ASI. NNyA.

Contexto

El programa “Las víctimas contra las violencias” fue creado en 2006 por el Ministerio de Justicia y Derechos Humanos de la Nación, para brindar atención a víctimas de abuso o malos tratos, de explotación y/o prostitución infantil. En conjunto con Unicef Argentina, entre octubre de 2020 y septiembre de 2021 se registraron 3.219 niñas, niños o adolescentes víctimas en consultas por violencia sexual. De ese total, un 77% eran niñas de 12 a 17 años [5]. Ante estos datos se han desarrollado diferentes iniciativas correspondientes a la prevención del abuso sexual infantil, tales como el programa de Educación Sexual Integral (ESI), Ley Nacional N° 26.150 [6]. También se han venido generando campañas preventivas concientizando y enseñando acerca de

los derechos de los niños y las responsabilidades de los adultos [10].

Introducción

Los videojuegos son un recurso muy valioso como método educativo y se los puede categorizar como *serious game* o juego serio [8].

Antecedentes relevantes se encuentran en casos como NOABS [9], un videojuego que se desarrolló con el fin de encontrar focos de vulnerabilidad en el jugador-niño. Actualmente este no se encuentra disponible para descarga y se debe ingresar con credenciales de inicio de sesión.

"Diana frente al espejo" [10] es un juego narrativo, creado en Málaga. Se desarrolla en primera persona, tiene por objetivo la prevención contra la violencia de género. Se presenta en un entorno virtual 3D interactivo, con cinemáticas y actores reales que representan personajes cuyas historias dependen de las decisiones que tome el jugador/a.

El videojuego, en su categoría de juego serio, puede fomentar el desarrollo de habilidades ejecutivas, recursos cognitivos fundamentales para la adaptación al medio de manera exitosa y la toma de decisiones [11].

Con base en estos argumentos se ideó un videojuego de índole preventivo que se enfoca en el aprender haciendo, además de inculcar conceptos teóricos e información para posibles denuncias. Esta forma de enseñanza está fundamentada por la Teoría del Aprendizaje Social de Albert Bandura [12]. Bandura sostenía que lo que observamos y lo que pensamos está relacionado con aquello que aprendemos y cómo lo aprendemos, propuso que la observación, la imitación y el modelaje son centrales para que se produzca el aprendizaje. El niño imitará o repetirá una conducta con resultado positivo y evitará aquella que deriva en consecuencias desagradables.

En base a esta teoría se ha decidido que el videojuego tenga mecánicas de elección de toma de decisiones, a sabiendas que habrá diálogos básicos que el niño-jugador deba leer y comprender. Considerando la variedad de capacidades lectoras, y de compresión de texto en estas edades [13] el diálogo entre personajes es acotado, aunque se refuerza con el empleo de imágenes o íconos que informan del tono.

Siendo así, este proyecto se centra en el diseño de un videojuego serio que aborda la prevención del Abuso Sexual Infantil de manera efectiva para niños y niñas de 10 y 11 años.

Líneas de Investigación, Desarrollo e Innovación

La propuesta se enmarca en un diseño de proyecto de intervención, a partir del cual, se espera realizar un aporte a la comunidad brindando herramientas para la prevención de una problemática social.

Como relevamiento inicial se realizó una encuesta a alumnos de 5to y 6to grado de escuelas primarias de gestión estatal con el objetivo de indagar los tipos de videojuegos que los niños y niñas juegan frecuentemente. Se buscó categorizar las características de estos juegos para identificar criterios a tener en cuenta en el desarrollo del videojuego.

Entre las características principales que se hallaron en común se destacan: Juegos en 3era persona, con capacidad de exploración de zonas, con gráficos llamativos pero no infantiles, personajes jóvenes adultos (no niños ni animales) y mecánicas diversas que requieren de poner en juego las capacidades atencionales.

Se desarrollaron diversas misiones con varias capas de mecánicas y objetivos para el jugador buscando

capturar y sostener la atención del jugador. Cada misión tiene su propia mecánica particular a forma de minijuego. Se estima que, con la baja repetición de estilos jugables durante el transcurso del videojuego, los niños-jugadores, no se desviarán de los eventos de la historia.

Además, las múltiples capas mecánicas que ofrecen cada misión requieren que el jugador aprenda nuevas situaciones de la vida cotidiana que les servirá para prevenir cualquier tipo de abuso. Entre estas se pueden mencionar la capacidad de mantener la atención en su entorno, mientras se realiza una tarea que requiera de cierto nivel de concentración.

Será necesario atender a la capacidad de escuchar diferentes puntos de vista, incorrectos, inmorales o ilegales, en donde el jugador deberá sostener su propio punto de vista sin ser alterado. Esta firmeza es crucial al distinguir entre actitudes que el niño considere aceptables y las que no para tener la capacidad de rechazar propuestas de otros y de dar su propia mirada sin que se vea distorsionada por los otros.

Estas misiones se crearán y desarrollarán en el motor de videojuegos Unity, que ya tuvo experiencia en su uso el equipo de investigadores.

Los modelados serán personalizados, utilizando herramientas como Blender para la construcción de las estructuras y de los personajes.

Se desarrollará un sistema de personalización de los personajes protagonistas, fomentando el interés del jugador a su personaje jugable.

En lo que refiere al diseño artístico, con base en los hallazgos de la encuesta, se ha elegido un personaje humano en 3D, con un estilo artístico flexible y llamativo. Se emplea una narrativa que explora emociones y épocas históricas, a fin de dar contraste entre lo ofrecido por los derechos de los niños y la realidad que se vivía antes de la existencia de estos.

El juego contará que el escenario se modificará según la época y la emoción del momento. Esta capacidad de transformación permitirá crear una experiencia visualmente dinámica y sorprendente, manteniendo a los jugadores interesados, pero también permitiendo representar situaciones de peligro.

Por ejemplo, el primer nivel combina el estilo románico clásico y el gótico los cuales se emplean para destacar lugares específicos, guiando al jugador por el mapa a distintos puntos de interés.

Este contraste de estilos artísticos y arquitectónicos ocurre con interés de generar una narrativa visual de corrupción en los espacios que se estará moviendo el jugador.

Poder generar la asociación de peligro a los puntos de interés donde ocurrirán las misiones podrá preparar al jugador para encontrarse con entornos hostiles. Además, se ofrece una barra de progreso simbólica donde el actuar del jugador con los protagonistas tiene peso y modifica su ambiente. Esta modificación se centra en eliminar estas corrupciones de tal modo de dejar el espacio en el que se encuentra el niño en un estado más natural y menos discordante.

Resultados y Objetivos

Los objetivos del proyecto son los siguientes:

1.Diseñar un videojuego para la promoción de la prevención del Abuso Sexual Infantil.

2.Fortalecer los factores protectores o, reducir los factores de riesgos de los NNyA [14], ante las situaciones de ASI. Se pretende que los usuarios de los juegos aprendan actitudes para prevenir exponerse a hechos victimizantes.

3.Crear una experiencia de juego inmersiva que enseñe a través de sus mecánicas, narrativa y arte.

4.Realizar pruebas en diferentes estadios del desarrollo del videojuego en ámbitos escolares.

5.Mejorar las mecánicas y el arte según los resultados de las pruebas anteriormente mencionadas.

Formación de Recursos Humanos

El proyecto está dirigido por la Dra. Cecilia Roma e integrado por la Lic. Julieta Otamendi, experta en grooming, y por los estudiantes Berenice Balich, Nahuel Buceta, Milena Manghi Y Scheck y Mateo Baya.

Referencias

[1] V. Berlinerblau, "Abuso sexual contra niños, niñas y adolescentes." UNICEF, 2017.

[2] O. Jaramillo y L. Castellón. "Educación y videojuegos." *Chasqui. Revista Latinoamericana de Comunicación*, no. 117, pp. 11-19, Febrero 2012.

[3] N. Geymonat Vignolo, "Videojuegos en seis escuelas de tiempo completo: puente entre lo sociocultural y lo didáctico pedagógico." *Cuadernos de investigación educativa*, vol.5, no. 20, pp. 11-19, Junio 2014.

[4] J. Sánchez Martínez (2018, marzo). Los videojuegos como herramienta de prevención y tratamiento de enfermedades. [Online].

[5] Ministerio de Justicia y Derechos Humanos Argentina. No 9. Serie violencia contra niñas, niños y adolescentes. Un análisis de los datos del Programa "Las Víctimas Contra Las Violencias" 2020-2021. UNICEF y el Programa Las Víctimas Contra Las Violencias.

[6] CFE. "Lineamientos Curriculares para la Educación Sexual Integral. Programa Nacional de Educación Sexual Integral." Ley Nacional N° 26.150 Res. CFE 340/18. Ministerio de Educación. Presidencia de la Nación.

[7] Ministerio de Educación de la Nación. "Educación Sexual Integral para charlar en familia", 2021. http://www.bnm.me.gov.ar/giga1/documentos/EL007593.pdf

[8] T. M. Fleming, et al.,"Serious Games and Gamification for Mental Health: Current Status and Promising Directions." *Frontiers in psychiatry*, no. 7, pp. 215, 2017. Serious Games and Gamification for Mental Health: Current Status and Promising Directions - PubMed

[9] R. Núñez, "NoAbs, un videojuego que busca detectar la vulnerabilidad de los niños de abuso sexual infantil." *Counter Pedophilia Investigative Unit: Contra la pedofilia online.* 2017

[10] Ministerio de Igualdad, Secretaría de Estado de Igualdad y contra la Violencia de Género. "Diana frente al espejo." Videojuego para la prevención de la violencia de género en jóvenes y adolescentes. Área de Igualdad del Ayuntamiento de Málaga, Cátedra Estratégica Interactividad y Diseño de Experiencias y Vicerrectorado de Igualdad de la Universidad de Málaga, 2020. Diana frente al espejo - DIANA

[11] C. Roma, "Relación entre videojuegos y funciones ejecutivas en niños de nivel primario: una revisión sistemática." *CAIS - Congreso Argentino de Informática y Salud,* vol. 10, no. 4, pp. 83.

[12]. C. Vergara Cano, (2024, mayo 8). "Bandura y la teoría del aprendizaje social." Actualidad en Psicología

[13] I. Kit, M. Nistal & E. Orlicki, "Los más favorecidos de Argentina, entre los menos favorecidos de la región. Resultados PISA 2022". Observatorio de Argentinos por la Educación, 2024. version final

[14] "Guía de intervención frente a abusos sexuales en niños niñas y adolescentes" https://argentina.unfpa.org/sites/default/files/pub-pdf/guia_de_intervencion_frente_a_abusos_sexuales_en_ninos_ninas_y_adolescentes.pdf

Trazando el Futuro Educativo: Exploración, Afinamiento y Aplicación de grandes modelos del lenguaje en la Educación Superior

Juan Pablo Tessore[1], Claudia Russo[2], Hugo Ramón[2]

Instituto de Investigación y Transferencia en Tecnología (ITT)
Centro asociado a la Comisión de Investigaciones Científicas (CIC)
Escuela de Tecnología (ET)
Universidad Nacional del Noroeste de la Provincia de Buenos Aires (UNNOBA)

Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)

juanpablo.tessore@itt.unnoba.edu.ar, claudia.russo@itt.unnoba.edu.ar, hugo.ramon@itt.unnoba.edu.ar

Resumen

Los grandes modelos de lenguaje (LLM) han transformado el procesamiento del lenguaje natural (NLP) mediante arquitecturas avanzadas como el Transformer y su mecanismo de autoatención. Modelos como GPT, LaMDA, LLaMA y BERT han impulsado aplicaciones en diversos sectores, incluyendo la educación. En el ámbito educativo, los LLM ofrecen oportunidades para personalizar el aprendizaje, adaptar contenido según las necesidades individuales de los estudiantes y mejorar la interacción mediante experiencias inmersivas. Además, facilitan el acceso a información relevante, optimizando el tiempo de docentes y administrativos al reducir tareas repetitivas.

Sin embargo, la integración de los LLM en la educación plantea desafíos significativos, como la privacidad de los datos, la equidad en el acceso tecnológico y la ética en el uso de la inteligencia artificial. Asimismo, el análisis de los datos generados por la interacción con estos modelos puede proporcionar información valiosa para mejorar los procesos de enseñanza, identificar patrones de aprendizaje y ofrecer retroalimentación en tiempo real.

Este trabajo busca evaluar el impacto de los LLM en la educación, explorando sus beneficios, limitaciones y riesgos. Además, de analizar las mejores prácticas para su implementación efectiva y ética en la comunidad universitaria, garantizando un uso responsable y equitativo.

***Palabras clave*:** Grandes modelos del lenguaje, Inteligencia artificial, Educación.

Contexto

Esta línea de investigación forma parte del proyecto "Inteligencia Artificial como herramienta para innovar y dinamizar procesos" presentado en el marco de la convocatoria a Subsidios de Investigación Bianuales (SIB 2022) de la Secretaría de Investigación, Desarrollo y Transferencia de

[1] Docente Investigador ITT - Becario Postdoctoral CONICET

[2] Docente Investigador ITT - - Investigador Asociado Adjunto sin director CIC

la UNNOBA. A su vez se enmarca en el contexto de un plan de trabajo aprobado por la Comisión Nacional de Investigaciones Científicas y Técnicas, en el marco de la convocatoria "Becas Postdoctorales 2022 Temas Estratégicos".

El proyecto se desarrolla en el Instituto de Investigación y Transferencia en Tecnología (ITT) dependiente de la mencionada Secretaría, y se trabaja en conjunto con la Escuela de Tecnología de la UNNOBA.

El equipo está constituido por docentes e investigadores pertenecientes al ITT y a otros Institutos de Investigación, así como también, estudiantes de las carreras de Informática de la Escuela de Tecnología de la UNNOBA.

Introducción

El origen de los grandes modelos de lenguaje (LLM, por sus siglas en inglés) comenzó con el desarrollo de redes neuronales para resolver tareas específicas relacionadas al procesamiento del lenguaje natural (NLP, por sus siglas en inglés). Partiendo desde la base de las redes neuronales recurrentes (RNN, por sus siglas en inglés) [1] hasta arquitecturas más complejas como las redes de Memorias de Corto y Largo Plazo (LSTM, por sus siglas en inglés) [2] y finalmente hasta el modelo Transformer [3] caracterizado por su mecanismo de autoatención, ha demostrado ser altamente efectivo en capturar detalles específicos en el lenguaje natural, permitiendo el desarrollo de los LLM.

A partir de allí, los LLM causaron una revolución en el NLP debido a su capacidad de interacción con el lenguaje humano. Modelos como el Transformador generativo pre entrenado (GPT por sus siglas en inglés) [4], el modelo del lenguaje para aplicaciones de dialogo (LaMDA, por sus siglas en inglés) [5], o el Gran modelo de Lenguaje Meta IA (LLaMA, por sus siglas en inglés) [6] basados en decoders, o la Representación de Codificador Bidireccional de Transformadores (BERT por sus siglas en inglés) [7] basado en encoders, y sus sucesores, marcaron hitos en sus campos de investigación permitiendo la aplicación de estas tecnologías en múltiples sectores como las finanzas, salud, comercio electrónico, educación, etc.

En el ámbito educativo, los LLM tienen el potencial de revolucionar la forma en que se enseña y se aprende, ofreciendo herramientas avanzadas para personalizar el aprendizaje, adaptar el contenido educativo según las necesidades individuales de los estudiantes, y crear experiencias educativas interactivas e inmersivas [8][9][10][11][12]. Cómo así también para acelerar el acceso a información relevante por parte de la comunidad universitaria [13][14][15]. Esto permitiría optimizar los recursos, ya que el personal (docente o administrativo) podría dedicar menor cantidad de tiempo en actividades rutinarias y/o repetitivas [16].

Por otro lado, la implementación de los LLM en la educación puede generar muchos datos producto de la interacción de las/os estudiantes con dichas tecnologías, lo que puede ser utilizado para ayudar a los educadores a identificar patrones de aprendizaje, adaptar el material didáctico de manera dinámica y ofrecer retroalimentación inmediata a los estudiantes, promoviendo así un aprendizaje más efectivo y autónomo.

Sin embargo, al implementar los LLM en entornos educativos, es crucial abordar cuestiones relacionadas con la privacidad de los datos, la equidad en el acceso a la tecnología y la ética en el uso de la

inteligencia artificial [18]. Se deben establecer políticas claras y directrices éticas para garantizar la protección de la información personal de los estudiantes y promover un uso responsable de la tecnología.

En este contexto, la presente línea de investigación tiene como propósito evaluar cada una de las interrogantes planteadas en los párrafos anteriores con respecto al potencial trasformador de los LLM en el ámbito educativo, el uso de los datos generados a partir de la interacción de los usuarios con estos modelos, la determinación de las mejores prácticas para la incorporación de los modelos en la comunidad educativa y el abordaje de cuestiones éticas y de privacidad de los datos.

Líneas de Investigación, Desarrollo e Innovación

La presente investigación se encuadra dentro del objetivo "Identificar, analizar, seleccionar e intervenir procesos utilizando herramientas y técnicas de la IA para favorecer su optimización" del mencionado proyecto SIB. Dicho objetivo incluye:

- El relevamiento actividades económicas primarias, secundarias, terciarias y cuartarias con potencial para ser mejoradas,
- El análisis y selección de aquellas actividades cuyos procesos presentan mayor potencial para ser mejorados, generando innovación a través de la IA.
- La identificación, adaptación, diseño y/o implementación de agentes inteligentes para generar ventajas competitivas en los procesos seleccionados.
- La determinación del impacto que estos agentes inteligentes provocan en los procesos intervenidos.

En tal sentido la investigación presente investigación acerca de los LLM en educación superior contribuye al proyecto SIB al abordar:

- El relevamiento y selección de actividades con potencial de mejora mediante IA, como la personalización del aprendizaje y la automatización de tareas administrativas.
- El diseño e implementación de agentes inteligentes que optimicen estos procesos, evaluando su impacto en la enseñanza y la gestión educativa.
- El análisis de aspectos éticos y de privacidad, asegurando una integración responsable y efectiva de la IA en el ámbito educativo.

Resultados y Objetivos

Se espera que esta investigación ofrezca una visión clara de cómo los grandes modelos de lenguaje (LLMs) pueden integrarse en el ámbito educativo universitario. A través de un análisis detallado de la literatura existente y pruebas con modelos pre entrenados, se busca comprender mejor lo que estos modelos pueden hacer, sus limitaciones y cómo se pueden aplicar en educación superior. Esta comprensión permitirá crear enfoques prácticos para su uso en las aulas, ayudando a optimizar tanto las tareas docentes como las administrativas.

Uno de los objetivos clave es identificar los usos más efectivos de los LLMs en educación. Entre los más prometedores se incluyen el apoyo en la enseñanza, la

generación automática de contenidos educativos y la personalización del aprendizaje según las necesidades de cada estudiante. También se documentarán los beneficios, así como los desafíos técnicos y éticos que surgen, proporcionando pautas claras para su adopción en las universidades.

Además, el trabajo dará lugar a una metodología detallada para implementar estos modelos en el entorno educativo. Esta metodología cubrirá aspectos como la selección de datos adecuados, las estrategias para adaptar los modelos a contextos específicos y cómo evaluar su desempeño. Con este enfoque, se espera mejorar la precisión de los modelos y asegurar que se alineen con los objetivos educativos y las necesidades de los estudiantes.

Finalmente, la investigación evaluará cómo el uso de los LLMs influye en el aprendizaje universitario, utilizando tanto datos cuantitativos como cualitativos. Los resultados ayudarán a entender mejor cómo estos modelos contribuyen a la generación de conocimiento, a la interacción entre estudiantes y docentes, y a la mejora de la eficiencia educativa.

Formación de Recursos Humanos

En esta línea de I/D se han obtenido y se encuentran desarrollando actualmente una beca postdoctoral. Asimismo, se esperan desarrollar una tesina de grado y/o PPS, dirigida por miembros de este proyecto.

Bibliografía

[1] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning Internal Representations by Error Propagation, Parallel Distributed Processing, Explorations in the Microstructure of Cognition, ed. DE Rumelhart and J. McClelland. Vol. 1. 1986. Biometrika, 71, 599-607.

[2] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.

[3] Shaw, P., Uszkoreit, J., & Vaswani, A. (2018). Self-attention with relative position representations. arXiv preprint arXiv:1803.02155.

[4] Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training.

[5] Thoppilan, R., De Freitas, D., Hall, J., Shazeer, N., Kulshreshtha, A., Cheng, H. T., ... & Le, Q. (2022). Lamda: Language models for dialog applications. arXiv preprint arXiv:2201.08239.

[6] Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M. A., Lacroix, T., ... & Lample, G. (2023). Llama: Open and efficient foundation language models. arXiv preprint arXiv:2302.13971. Fecha 30/04/2024 Página 9

[7] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

[8] Zhao, W.X.; Zhou, K.; Li, J.; Tang, T.; Wang, X.; Hou, Y.; Min, Y.; Zhang, B.; Zhang, J.; Dong, Z.; et al. A survey of large language models. arXiv 2023, arXiv:2303.18223.

[9] Liu, Y.; Han, T.; Ma, S.; Zhang, J.; Yang, Y.; Tian, J.; He, H.; Li, A.; He, M.; Liu, Z.; et al. Summary of ChatGPT/GPT-4 research and perspective towards the future of large language models. arXiv 2023, arXiv:2304.01852.

[10] Rasul, T., Nair, S., Kalendra, D., Robin, M., de Oliveira Santini, F., Ladeira, W. J., ... & Heathcote, L. (2023). The role of ChatGPT in higher education: Benefits, challenges, and

future research directions. Journal of Applied Learning and Teaching, 6(1).

[11] Atlas, S. (2023). ChatGPT for higher education and professional development: A guide to conversational AI.

[12] Gill, S. S., Xu, M., Patros, P., Wu, H., Kaur, R., Kaur, K., ... & Buyya, R. (2024). Transformative effects of ChatGPT on modern education: Emerging Era of AI Chatbots. Internet of Things and Cyber-Physical Systems, 4, 19-23.

[13] Julius Odede and Ingo Frommholz. 2024. JayBot – Aiding University Students and Admission with an LLM-based Chatbot. In Proceedings of ACM SIGIR Conference on Human Information Interaction And Retrieval (CHIIR '24). ACM, New York, NY, USA, 5 pages.

[14] Bratić, D., Šapina, M., Jurečić, D., & Žiljak Gršić, J. (2024). Centralized Database Access: Transformer Framework and LLM/Chatbot Integration-Based Hybrid Model. Applied System Innovation, 7(1), 17.

[15] Oliveira, P. F., & Matos, P. (2023). Introducing a Chatbot to the Web Portal of a Higher Education Institution to Enhance Student Interaction. Engineering Proceedings, 56(1), 128.

[16] Deng, X., & Yu, Z. (2023). A meta-analysis and systematic review of the effect of chatbot technology use in sustainable education. Sustainability, 15(4), 2940.

[17] Rudolph, J., Tan, S., & Tan, S. (2023). War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. Journal of Applied Learning and Teaching, 6(1).

[18] Dempere J, Modugu K, Hesham A and Ramasamy LK (2023) The impact of ChatGPT on higher education. Front. Educ. 8:1206936. doi: 10.3389/feduc.2023.1206936

Un abordaje de la enseñanza de algoritmos evolutivos en la materia inteligencia artificial en la carrera de ISI UTN FRC desde los ambientes de aprendizaje colaborativos

Olariaga, Sandra [1]; Paéz, Nancy del Valle [2]; Stefanich, Clarisa [3];
Hass, Silvina [4]; Giordano, Franco [5]
Departamento de Ingeniería en Sistemas de Información
Universidad Tecnológica Nacional, Facultad Regional Córdoba (UTN-FRC)
Maestro Marcelo López y Cruz Roja Argentina – Ciudad Universitaria – Córdoba, Argentina
(1)solariaga@yahoo.com; (2) nancydelv@gmail.com ; (3) clarystefanich@gmail.com
(4) silhass13@gmail.com ; (5) fgiordano@hotmail.com

Eje Temático del Trabajo: TIAE Tecnología Informática Aplicada en Educación.

RESUMEN

El uso de la tecnología como apoyo al proceso de enseñanza y aprendizaje, se ha convertido en necesario e imprescindible. Las estrategias didácticas mediadas por tecnologías permiten presentar conceptos e ideas a los estudiantes de manera multimedial, ágil e interactiva, situándolos en un rol activo y participativo, sumado a que se desarrollan ambientes virtuales que les permiten ejercitar y extender sus conocimientos. En el caso concreto de la asignatura Inteligencia Artificial, se vienen incorporando, herramientas que ayudan al estudio de los diferentes contenidos curriculares; en esta ocasión, el enfoque está puesto en la temática de algoritmos evolutivos. Con el nombre de algoritmos evolutivos se identifica a un grupo de procedimientos que se distinguen de otras técnicas porque están inspirados en las teorías que sintetizan la evolución Darwiniana y la herencia genética descripta por Mendel, también se basan en la inteligencia colectiva de grupos de organismos, como bandadas de aves o cardúmenes de peces, para resolver problemas de optimización.

A través de este proyecto se pretende focalizar en el uso de estas tecnologías como medio para que el estudiante construya su propio conocimiento, pero en una dinámica grupal y colaborativa, lo que actualmente se denomina ambiente colaborativo de aprendizaje

Palabras clave: algoritmos evolutivos, inteligencia artificial, aprendizaje colaborativo.

CONTEXTO

El presente proyecto está siendo llevado a cabo en el Departamento de Ingeniería en Sistemas de la Universidad Tecnológica Nacional Facultad Regional Córdoba. (U.T.N.-F.R.C.).

El Código del proyecto es TETEUCO0008805 dentro del área de Aplicaciones tecnológicas en la educación superior, radicado dentro de la Secretaría de Ciencia y Tecnología de la

Universidad Tecnológica Nacional Facultad Regional Córdoba. Fecha de inicio: 01/04/2023 Fecha de finalización: 31/12/2026.

1. INTRODUCCIÓN

La incorporación de las TIC a la sociedad, y en particular a la educación, generaron importantes cambios en la interacción y comunicación entre los estudiantes, el aprendizaje colaborativo fue una evidencia de este cambio, quien progresivamente ha ido evolucionando y abarcando todos los niveles de educación.

Los ambientes colaborativos de aprendizaje buscan propiciar espacios en los que se exploren los conceptos o las situaciones problemáticas que se desean tratar, apoyados con tecnología, así como de estrategias pedagógicas como la investigación y pensamiento crítico, de esta manera se busca impulsar el desarrollo de habilidades mixtas (aprendizaje y desarrollo personal y social), donde cada miembro del grupo es responsable tanto de su aprendizaje como del aprendizaje de los restantes miembros del grupo, lo que implica un cambio de roles asociados a profesores y estudiantes.

Este modelo permite aprovechar las ventajas del trabajo grupal y la colaboración entre los mismos, ya que cada estudiante asume su propio ritmo y pone en juego sus potencialidades, alentando la actividad autónoma, pero de manera colaborativa donde cada uno comprende la necesidad de aportar lo mejor de si al grupo, para lograr un resultado sinérgico; se logra así, una relación de interdependencia favoreciendo los procesos de aprendizaje individuales, las relaciones interpersonales y la productividad. Además, favorece la capacidad de resolver problemas de forma creativa, a partir de estrategias de negociación y mediación y la búsqueda cooperativa de alternativas, fomentando la motivación y una actitud positiva hacia la materia de estudio, hacia la figura del profesor y hacia la función de la institución formativa.

Con el nombre de algoritmos evolutivos se identifica a un grupo de procedimientos que se utilizan para resolver muchos problemas prácticos y que se distinguen de otras técnicas porque están inspirados en las teorías que sintetizan la evolución Darwiniana y la herencia genética descripta por Mendel. Estos algoritmos se destacan por combinar exitosamente dos mecanismos de búsqueda: exploración y explotación. La exploración permite identificar áreas prometedoras del espacio de búsqueda y la explotación lleva a cabo una búsqueda especializada dentro de estas áreas para encontrar soluciones cercanas al óptimo. Su aplicación se extiende en prácticamente todas las áreas de la actividad humana como por ejemplo en salud,

educación, ciencias, industria, economía y defensa. También incluimos a la optimización por enjambre de partículas (*Particle Swarm Optimization, PSO*) que es un método de optimización heurística orientado a encontrar mínimos o máximos globales. Su funcionamiento está inspirado en el comportamiento que tienen las bandadas de pájaros o bancos de peces en los que, el movimiento de cada individuo (dirección, velocidad, aceleración) es el resultado de combinar las decisiones individuales de cada uno con el comportamiento del resto.

La estructura de un algoritmo *PSO* para optimizar (maximizar o minimizar) una función con una o múltiples variables sigue los siguientes pasos:

1 Crear un enjambre inicial de n partículas aleatorias. Cada partícula consta de 4 elementos: **una posición** que representa una determinada combinación de valores de las variables, **el valor de la función objetivo** en la posición donde se encuentra la partícula, **una velocidad** que indica cómo y hacia donde se desplaza la partícula, y **un registro de la mejor posición** en la que ha estado la partícula hasta el momento.

2 Evaluar cada partícula con la función objetivo.

3 Actualizar la posición y velocidad de cada partícula. Aquí es donde reside la capacidad de optimización del algoritmo. Si no se cumple un criterio de parada, volver al paso 2.

Las ventajas del uso de PSO son: menos parámetros a ajustar, es más fácil de implementar, mayor velocidad de convergencia y un uso eficiente de la memoria.

Bibliografía más importante consultada:

ADELL, J., Mengual-Andrés, S., & Roig-Vila, R. (2015). Presentación del Monográfico. Webquest: 20 años utilizando Internet como recurso para el aula. Edutec.

COELLO COELLO Carlos Artemio Coordinador Computación Evolutiva en colaboración con la Academia Mexicana de Computación Primera edición: 2019

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El proyecto se inscribe dentro de los lineamientos de investigación en innovaciones curriculares en Educación Superior Universitaria dentro del área de ambientes colaborativos de aprendizaje.

Para el desarrollo de este trabajo de investigación se aplicará el método empírico-analítico, que se basa en la experimentación y en la lógica empírica, junto a la aplicación y evaluación de la herramienta elaborada.

3. RESULTADOS OBTENIDOS / ESPERADOS

El eje sobre el que gira nuestro trabajo es la elaboración e implementación de una

herramienta de software que nos permite evaluar la enseñanza de la temática de algoritmos evolutivos.

Dicha evaluación está basada en el análisis de los datos recabados cuando se llegue a implementar la herramienta desarrollada.

Los docentes son los actores que mayores cambios pueden generar en las universidades ya que en ellos recae la responsabilidad de la formación y la orientación de los alumnos.

Los resultados obtenidos serán presentados ante el jefe de cátedra de la materia Inteligencia Artificial, para que previa evaluación pueda ser implementados en todos los cursos que integran dicha cátedra, a través de la aplicación de una herramienta que aprovecha las ventajas del trabajo grupal y colaborativo.

La implementación de esta herramienta didáctica permitirá crear una actitud autogestionada por parte del estudiante para la obtención y transformación de la información en nuevos conocimientos en pos de una mejora continua del proceso enseñanza aprendizaje y realizando los procesos característicos de la profesión de un ingeniero en sistemas.

Este proyecto es, por lo tanto, un aporte para la difusión de nuevas formas, aproximaciones y aspectos particulares en la implementación de las nuevas tecnologías educativas, en el ámbito. universitario, propiciando estrategias didácticas innovadoras

4. FORMACIÓN DE RECURSOS HUMANOS

El desarrollo del Proyecto que engloba este trabajo, contribuirá a ampliar el conocimiento acerca de la aplicación de ambientes colaborativos de aprendizaje, de manera de que los resultados obtenidos puedan llegar a ser el punto de partida de actividades que favorezcan el mejoramiento de los procesos de enseñanza y de aprendizaje en el aula universitaria.

Los investigadores que recién inician forjarán sus primeras herramientas en I+D lo que les permitirá ampliar conocimientos en el área en cuestión además de incrementar sus antecedentes académicos. Este proyecto contribuirá a fortalecer la formación de un becario, estudiante de Ingeniería en Sistemas de Información.

5. BIBLIOGRAFÍA

ADELL, J., Mengual-Andrés, S., & Roig-Vila, R. (2015). Presentación del Monográfico. Webquest: 20 años utilizando Internet como recurso para el aula. Edutec. Revista Electrónica de Tecnología Educativa, 52. http://www.edutec.es/revista/index.php/edutec-e/article/view/622

COELLO Carlos Artemio Coordinador Computación Evolutiva en colaboración con la Academia Mexicana de Computación Primera edición: 2019 Academia Mexicana de Computación, A. C.

DÍAZ BARRIGA: Estrategias docentes para un aprendizaje significativo. Una interpretación constructivista. México, Mcgraw-Hill. (1999)

FERNANDEZ MUÑOZ, R. (2003). Competencias profesionales del docente en la sociedad del siglo XXI. Praxis, 1, 4-8.

GUTIERREZ ESTEBAN, P., Yuste, R., Cubo, S., & Lucero, M. (2011). Buenas prácticas en el desarrollo de trabajo colaborativo en materias TIC aplicadas a la educación. Profesorado, Revista de Currículum y Formación del Profesorado, 15. Recuperado de http://www.ugr.es/~recfpro/rev151ART13.pdf

LAISEMA, S., & Wannapiroon, P. (2014). Design of Collaborative Learning with Creative Problem-solving Process Learning Activities in a Ubiquitous Learning Environment to Develop Creative Thinking Skills. Procedia - Social and Behavioral Sciences, 116, 3921-3926. doi:10.1016/j.sbspro.2014.01.867

LUCERO, M. M: Entre el trabajo colaborativo y el aprendizaje colaborativo Revista Iberoamericana de Educación ISSN 1681-5653 (2003).

MORIN, E. Los siete saberes necesarios para la educación del futuro. Organización de las Naciones Unidas, para la Educación, la Ciencia y la Cultura.Traducción de Mercedes Vallejo-Gómez, Medellín. Colombia. (2000).

SALINAS, J., de Benito, B., & Lizana, A. (2014). Competencias docentes para los nuevos escenarios de aprendizaje. Revista Interuniversitaria de Formación del Profesorado, 28(1), 145-163. Recuperado de http://www.redalyc.org/articulo.oa?id=27431190010.

TUDGE: Vigotsky: la zona de desarrollo próximo y su colaboración en la práctica de aula. NuevaYork, Universidad de Cambridge. (1994).

VIGOTSKY, L. (1979). El desarrollo de los procesos psicológicos superiores.

ZHU, C. (2012). Student Satisfaction, Performance, and Knowledge Construction in Online Collaborative Learning. Educational Technology & Society, 15(1), 127-136.

Uso de Analítica Predictiva para Anticipar el Desempeño Académico en Materias del Área Programación

Carrizo Claudio, Javier Saldarini, Raúl Navarro Peláez, Daiana Viotti, Antonella Calloni
Facultad Regional San Francisco – Universidad Tecnológica Nacional
Av. de la Universidad 501 - San Francisco - Córdoba - Tel. 03564-421147
{cjcarrizo77, saldarinijavier, ingrnavarropelaez, dayviotti2015, antocalloni}@gmail.com

RESUMEN

Hoy en día es fundamental que las universidades hagan uso de la analítica de datos, con el fin de generar conocimiento que sea de utilidad, para apoyar el proceso de toma de decisiones. Uno de los principales desafíos consiste en mantener un alto nivel de rendimiento académico, a fin de evitar la deserción por parte de los estudiantes de carreras de grado. En el ámbito de una Universidad Púbica, se detectaron problemas de rendimiento académico en instancias de evaluación parcial en materias del área de programación del primer nivel, esto provoca en muchos casos que los estudiantes abandonen el cursado de dichas materias.

En este sentido, es posible aplicar analítica predictiva con el propósito de identificar patrones en el conjunto de datos a analizar, con el fin de determinar en forma temprana el desempeño académico que tendrán los futuros estudiantes al cursar estas materias; esto permitirá generar una alerta en pos de definir acciones en forma temprana, que permitan mejorar el rendimiento estudiantil.

En la presente línea de investigación se propone la construcción de un modelo predictivo que permita obtener información de antemano para definir la ejecución de un conjunto de acciones, en pos de mejorar el proceso educativo.

Palabras clave: analítica académica, analítica predictiva, desempeño académico, área programación

CONTEXTO

La línea de investigación se enmarca dentro del Proyecto de Investigación y Desarrollo (PID) denominado “Analítica Predictiva aplicada para la mejora del desempeño académico en UTN Facultad Regional San Francisco”, el cual fue homologado por la Secretaría de Ciencia, Tecnología y Posgrado de la Universidad Tecnológica Nacional, bajo el código TEECSF0010247 y según la Disp. SCTyP Nº 32-2023. Cabe destacar que este PID se viene ejecutando desde inicios del año 2024.

1. INTRODUCCIÓN

Las organizaciones generan en la actualidad grandes volúmenes de datos, qué por sí solos, no aportan valor (Adi & Isnanto, 2020). Aquellas organizaciones que no transformen los datos e información en conocimiento, estarán desperdiciando uno de sus activos más importantes (Sumpala & Rasyid, 2019); de lo anterior surge la importancia de hacer uso de la analítica de datos.

Esta disciplina busca no solo brindar información objetiva y precisa para soportar la toma de decisiones

estratégicas, sino también mantener la competitividad de las organizaciones, mediante la aplicación de modelos analíticos y predictivos, que van más allá de las estadísticas descriptivas.

En este contexto, la educación superior no es la excepción, ya que uno de sus principales retos es obtener conocimiento, el cual permita apoyar el proceso de toma de decisiones, y generar una especie de inteligencia que derive en la definición de un conjunto de acciones en pos de la mejora de la gestión. Las universidades deben hacer uso de la analítica, como una nueva oportunidad para cumplir con esta responsabilidad, apoyados en las Tecnologías de la Información y Comunicación (TIC); esto permitirá poner en foco sus procesos y herramientas aplicadas en la toma de decisiones (Saiful & Nur, 2020).

El térmico "analítica" involucra diversas técnicas, herramientas y métodos que derivan de diversas disciplinas como son la estadística (Baepler & Murdoch, 2010b), inteligencia de negocios (Argonza, 2016), minería de datos (García-González et al., 2019) y el aprendizaje automático (Burman & Som, 2019).

En el contexto de la educación, existen también 3 tipos de analítica (Murnion y Helfert, 2013), las cuales se detallan a continuación:

1. Analítica descriptiva: se enfoca en el desempeño del total de las actividades.

2. Analítica diagnóstica: se enfoca en el desempeño pasado para analizar la información.

3. Analítica predictiva: se anticipa a los comportamientos basándose en relaciones históricas entre las variables.

Por otra parte, existen también 2 grandes enfoques en el contexto del uso de la analítica en la educación superior:

1. Analítica Académica: consiste en diversos procesos de evaluación y análisis de datos que permiten medir, mejorar y comparar el desempeño de individuos, programas o instituciones. Es un análisis enfocado en aspectos académicos, relativos al éxito de los estudiantes según (Norris & Lefrere, 2011).

2. Analítica del Aprendizaje: consiste en la medición, recopilación, análisis y presentación de los datos sobre los estudiantes, sus contextos y las interacciones que se generan (Siemens y Gasevi, 2012) (Amo y Santiago, 2017). Se centra en la comprensión, estudio y predicción del aprendizaje de los estudiantes apoyados en el estudio virtual (Viberg et al., 2018).

La analítica académica parte de la recopilación de la información y su análisis por medio de la minería de datos, la cual posee diversos métodos de aplicación; en algunos de ellos se trata de la búsqueda de patrones sobre el conjunto de datos que se desea analizar (Daniel, 2015a).

Uno de los aspectos en donde las universidades deben poner el foco es en el monitoreo del rendimiento académico, ya que comúnmente se asocia un bajo rendimiento académico, con una alta tasa de deserción. El rendimiento académico se puede definir como la productividad del sujeto, matizado por sus actividades, rasgos y la percepción más o menos correcta de los cometidos asignados (Maletic et. al., 2002). Generalmente el rendimiento académico se evalúa teniendo en cuenta aspectos que influyen en el desempeño, pudiendo mencionar entre ellos: factores socioeconómicos, programas de estudio, metodologías de

enseñanza, conocimientos previos del alumno.

En (Gómez, 2015) se conceptúa que la medición del rendimiento educativo por medio de la analítica incluye dos aspectos importantes que se complementan. Por un lado, debe considerarse el rendimiento operativo (sistemas administrativos y de apoyo), y por el otro, el rendimiento académico (diseño y ejecución de estrategias académicas para lograr experiencias de aprendizaje, los resultados y las competencias de la vida real).

Para evaluar el rendimiento académico de un alumno es necesario conocer si existen perfiles comunes a grupos de estudiantes, de esta manera, la determinación de estos perfiles se convierte en una estrategia de valor significativo a la hora de tomar acciones, que permitan mejorar el desempeño de los alumnos (Di Gresia, 2007).

El establecimiento de perfiles consiste en el proceso de determinación y clasificación de patrones; en este sentido se hace uso de la minería de datos educacional, la cual consiste en la aplicación y uso de métodos que permiten explorar datos que provienen de ambientes educativos, en pos de la mejora en la comprensión de los estudiantes y el entorno en el que aprenden (International Educational Data Mining Society).

Según datos estadísticos brindados por el Área de Gestión Académica de una Facultad de Ingeniería de una Universidad Pública, existe una alta tasa de estudiantes con bajo rendimiento académico en instancias de evaluación parcial en materias del Área de Programación del primer nivel de una carrera de Ingeniería; esto provoca que muchos de ellos no alcancen la regularidad, por lo que esto deriva en muchos casos en el abandono del cursado de dichas cátedras.

Visto el marco descripto anteriormente, nos impulsó a iniciar una nueva línea de investigación, cuyo objetivo principal es hacer uso de la analítica académica para generar un modelo que permita predecir el desempeño de estudiantes en instancias de evaluación parcial en materias del área de programación del primer nivel; en este sentido, se construirá el perfil del alumno que tenga bajo rendimiento en instancias de evaluación parcial.

El principal aporte de esta línea de investigación consiste en poder obtener información de antemano, la cual sirva de instrumento para ejecutar acciones en forma anticipada, en pos de mejorar el proceso educativo, evitando la deserción en el cursado de dichas materias.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Los ejes de trabajo de la línea se investigación, se detallan a continuación:

- Elaboración del perfil del estudiante de primer año.
- Relevamiento, descripción y exploración de los datos, según perfil del estudiante.
- Pre-procesamiento de los datos relevados (limpieza, transformación y formateo).
- Análisis exploratorio de datos mediante estadística descriptiva.
- Selección del conjunto de atributos aptos para el modelo predictivo.
- Estudio en profundidad de los algoritmos de Machine Learning para aprendizaje supervisado.
- Elaboración de modelos predictivos, utilizando distintos tipos de algoritmos.
- Validación de los modelos predictivos.
- Elección del modelo predictivo que brinde los mejores resultados.
- Visualización de los resultados obtenidos en alguna herramienta afín.
- Comunicación de resultados obtenidos al Área de Gestión Académica y docentes de materias del Área de Programación.

3. RESULTADOS

Resultados Obtenidos

- Se logró construir el perfil del estudiante de primer año, compuesto por datos personales, geográficos, académicos y socio-económicos.
- Se obtuvo un set de datos de cursado de estudiantes en materias del área de programación, compuesto por 33 atributos y 2335 registros.
- Se realizó una descripción de cada uno de los atributos del set de datos.
- Se realizó una exploración de datos inicial, a través de una herramienta de visualización, a fin de poder realizar un análisis preliminar de los datos de algunos atributos incluidos en el dataset.
- Se realizaron tareas de pre-procesamiento de datos usando el lenguaje de programación Python, específicamente relacionadas con limpieza, transformación y calidad de datos.
- Se realizó un análisis exploratorio de los datos, utilizando estadística descriptiva.
- Se realizó la selección del conjunto de atributos que serán incluidos en el modelo predictivo.

Resultados Esperados

- Informe que contenga el resultado del estudio de algoritmos de aprendizaje supervisado.
- Selección de los algoritmos más apropiados para la solución propuesta.
- Elaborar y validar cada uno de modelos predictivos.
- Evaluar el resultado de cada modelo.
- Comunicar los resultados al Área Académica y docentes través de una herramienta de visualización.
- Transferir los resultados de la investigación al ámbito académico.
- Publicar los resultados de la investigación en algún congreso científico.

FORMACIÓN DE RECURSOS HUMANOS

Esta línea de investigación se trabaja desde el Grupo de Investigación y Desarrollo de UTN Facultad Regional San Francisco.

El equipo de trabajo está constituido por: 1 Director de Proyecto, 1 Co-Director, 3 docentes investigadores de apoyo, 2 estudiantes de nivel avanzado, todos de la especialidad de Ingeniería en Sistemas de Información.

Cabe destacar, que este equipo de trabajo tiene antecedentes de haber desarrollado y obtenido resultados en PID de UTN, vinculados a la temática “Minería de datos educacional”. También se han presentado resultados en jornadas, workshop y congresos científicos de la especialidad.

5. BIBLIOGRAFÍA

Adi, K., & Isnanto, R. R. (2020). Rice crop management expert system with forwarding chaining method and certainty factor. In Journal of Physics: Conference Series (Vol. 1524, No. 1, p. 012037). IOP Publishing. https://iopscience.iop.org/article/10.1088/1742-6596/1524/1/012037/pdf

Amo, D. y Santiago, R. (2017) Learning Analytics. La narración del aprendizaje a través de los datos. Barcelona: Editorial UOC. Pixel-Bit.

Argonza, J. S. (2016). Big data en la educación. Vol. 17, No. 1.

Baepler, P., & Murdoch, C. J. (2010). Academic analytics and data mining in higher education. International Journal for

the Scholarship of Teaching and Learning, 4(2), 17.

Burman, I., & Som, S. (2019, February). Predicting students academic performance using support vector machine. In 2019 Amity international conference on artificial intelligence (AICAI) (pp. 756-759). IEEE.

Daniel, B. (2015a). Big Data and analytics in higher education: Opportunities and challenges. British Journal of Educational Technology, 46–5(0007–1013), 904–920. https://www.learntechlib.org/p/151804/

Gómez, D. (2015). Analítica visual en eLearning [Universdiad de Salamanca]. https://gredos.usal.es/handle/10366/125950.

International Educational Data Mining Society. Disponible en https://uia.org/s/or/en/1122278867

Maldonado Berea, G. A., García González, J., & Sampedro Requena, B. E. (2019). El efecto de las TIC y redes sociales en estudiantes universitarios. RIED. Revista Iberoamericana de Educación a Distancia.

Maletic, J., Collard, M. & Marcus, A (2002). Source Code Files as Structured Documents. Proceedings 10th IEEE International Workshop on Program Comprehension (IWPC'02), 289- 292.

Marcus, A. (2003). Semantic Driven Program Analysis, Kent, OH, USA, Kent State University Doctoral Thesis.

Murnion, P., Helfert, M. (2013). Academic Analytics in Quality Assurance Using Organisational Analytical Capabilities. En UK Academy for Information Systems Conference Proceedings 2013.

Norris, D. M., & Lefrere, P. (2011). Transformation through expeditionary change using online learning and competencebuilding technologies. Research in Learning Technology, 19(1), 1741–1629. https://doi.org/10.3402/rlt.v19i1.17105

Saiful, M., & Nur, A. M. (2020). Application of Expert System with Web-Based Forward Chaining Method in Diagnosing Corn Plant Disease. In Journal of Physics: Conference Series, 1539(1)

Siemens, G., & Gaševi?, D. (2012). Special Issue on Learning and Knowledge Analytics. Educational Technology & Society, 15(3), 1–163.

Sumpala, A. T., & Rasyid, R. (2019). Expert system for corn plant disease diagnosis with the breadth-first search method. In IOP Conference Series: Earth and Environmental Science 382,(1) p. 012001). IOP Publishing. https://iopscience.iop.org/article/10.1088/1755-1315/382/1/012001/pdf

Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The current landscape of learning analytics in higher education. Computers in Human Behavior, 89(October 2017), 98–110. https://doi.org/10.1016/j.chb.2018.07.027



Tesis Doctorales

Análisis de trayectorias utilizando técnicas de minería de datos

Autor: Gary Reyes

Directora: Dra. Laura Lanzarini

Codirector: Dr. Waldo Hasperué

Tesis presentada para obtener el grado de Doctor en Ciencias Informáticas

Doctorado en Ciencias Informáticas, Facultad de Informática, Universidad Nacional de La Plata (UNLP)

Fecha de exposición: 12/12/2024

Síntesis

Motivación

El crecimiento acelerado del parque automotor en las ciudades ha generado problemas significativos de tráfico y congestión, afectando el tiempo de desplazamiento, la calidad del aire y la productividad económica. La gestión eficiente del tráfico requiere un análisis detallado no solo de la cantidad de vehículos en circulación, sino también de sus patrones de movimiento a lo largo del tiempo y el espacio. En este contexto, el estudio de trayectorias vehiculares se ha convertido en una herramienta fundamental para comprender la movilidad urbana y anticipar posibles escenarios de congestión.

El análisis de trayectorias vehiculares se basa en datos recolectados en tiempo real mediante dispositivos GPS, sensores y cámaras, generando un volumen masivo de información que exige técnicas avanzadas de procesamiento. La aplicación de Big Data y Machine Learning permite extraer información relevante de estos datos y mejorar la capacidad de toma de decisiones en la gestión del tráfico. Las técnicas de Machine Learning desempeñan un papel crucial en la identificación de congestiones, ya que pueden detectar patrones ocultos, predecir puntos críticos de tráfico y optimizar el flujo vehicular mediante ajustes en la infraestructura, como la sincronización de semáforos o la recomendación de desvíos estratégicos.

El uso de técnicas de minería de datos, en particular el agrupamiento dinámico, ha demostrado ser una estrategia efectiva para segmentar los flujos vehiculares en grupos homogéneos, lo que permite la identificación de zonas con alta densidad vehicular y la predicción de posibles embotellamientos. En este sentido, el flujo de datos es similar al flujo vehicular, ya que ambos presentan variaciones constantes y requieren métodos eficientes para su procesamiento y análisis en tiempo real. Algoritmos de agrupamiento como T-Optics y Tra-DBSCAN han sido empleados para detectar patrones recurrentes en el tráfico, facilitando la planificación de mejoras en la infraestructura vial y el diseño de redes de transporte público más eficientes.

Sin embargo, a pesar de los avances en estas técnicas, el agrupamiento de trayectorias vehiculares enfrenta varios desafíos que dificultan su implementación efectiva. Entre los problemas más relevantes se encuentran:

Variabilidad en los datos de trayectorias: Las diferencias en la longitud y frecuencia de muestreo de las trayectorias pueden afectar la generalización de los resultados y la identificación precisa de patrones.

Variabilidad a lo largo del tiempo: Los patrones de movimiento cambian dinámicamente, lo que dificulta la adaptación de algunos algoritmos tradicionales a estas fluctuaciones.

Visualización de resultados: La interpretación de los grupos generados por los algoritmos de agrupamiento sigue siendo un reto significativo, ya que los datos de trayectorias son multidimensionales y requieren herramientas especializadas para su análisis visual.

Uso de datos históricos: La disponibilidad de información sobre trayectorias pasadas es fundamental para identificar tendencias a largo plazo y mejorar la capacidad predictiva de los modelos propuestos.

Además de estos desafíos, otros factores como el ruido en los datos, la alta dimensionalidad de la información y los elevados requisitos computacionales también influyen en la efectividad de los métodos de agrupamiento. A pesar de estas dificultades, la integración de enfoques basados en Inteligencia Artificial y aprendizaje automático ofrece soluciones prometedoras para abordar estos problemas, permitiendo una adaptación más eficiente a las condiciones dinámicas del tráfico urbano.

En este contexto, la presente investigación propone el desarrollo de un método de agrupamiento dinámico de flujos de datos aplicado al análisis de trayectorias vehiculares. Este enfoque busca superar las limitaciones de los algoritmos tradicionales, proporcionando una solución adaptable y en tiempo real que permita la identificación eficiente de zonas con alta densidad vehicular. La combinación de técnicas avanzadas de procesamiento, modelos de aprendizaje automático y estrategias de visualización de datos contribuirá a mejorar la gestión del tráfico y a optimizar la movilidad en entornos urbanos.

Aporte de la tesis a la disciplina

El presente trabajo introduce un método innovador para la identificación de la variabilidad en el flujo vehicular, diseñado para operar en entornos urbanos con condiciones de tráfico cambiantes. La novedad del método radica en su capacidad para realizar procesamiento de flujos de datos en tiempo real, combinando dos niveles de agrupamiento que permiten detectar y analizar patrones de densidad vehicular. Este enfoque proporciona una solución efectiva para la gestión del tráfico y la identificación de zonas con posibles problemas de congestionamiento.

Componentes del método propuesto

El método se estructura en dos módulos principales y dos sub-módulos complementarios, que en conjunto garantizan la precisión del análisis de la variabilidad del flujo vehicular.

Módulo principal de nivel 1: Agrupamiento dinámico de ubicaciones

El primer nivel del método se centra en la formación de microgrupos, unidades de análisis conformadas por un conjunto de ubicaciones GPS. Cada microgrupo se define mediante un centroide representativo y un área de cobertura delimitada, facilitando la segmentación espacial del tráfico en zonas de comportamiento homogéneo.

El proceso de agrupamiento se desarrolla en tiempo real, mediante la evaluación continua de las ubicaciones GPS recibidas. Se implementa un criterio de similitud basado en la distancia euclidiana, determinando si una nueva ubicación GPS debe agregarse a un microgrupo existente o si es necesario crear uno nuevo. Además, se introduce un mecanismo de olvido porcentual de datos, que regula la permanencia de las ubicaciones en los microgrupos según su antigüedad, garantizando la representatividad actualizada de los datos.

Este mecanismo de olvido se aplica en dos niveles: en cada nueva actualización de un microgrupo y en intervalos predefinidos para eliminar microgrupos inactivos. De esta manera, el sistema evita la acumulación de datos obsoletos y se mantiene eficiente en su procesamiento.

Adicionalmente, cada 60 segundos se generan instantáneas del estado de los microgrupos formados, permitiendo capturar la evolución del tráfico a lo largo del tiempo. Estas instantáneas son insumos fundamentales para el segundo nivel de agrupamiento.

Módulo principal de nivel 2: Agrupamiento por proximidad

El segundo nivel del método se encarga de agrupar los microgrupos generados en el primer nivel según su proximidad, proporcionando una perspectiva agregada del tráfico. Este módulo no opera en tiempo real, sino que se ejecuta periódicamente sobre las instantáneas almacenadas, permitiendo identificar la evolución de las densidades vehiculares en diferentes momentos del día.

El proceso de agrupamiento por proximidad se desarrolla en tres fases:

Identificación del área preprocesada más cercana: Se determina la similitud entre los microgrupos actuales y las áreas preprocesadas con datos históricos, utilizando distancia euclidiana para optimizar la asignación de densidades referenciales.

Cálculo de densidad actual y clasificación de variabilidad: Se compara la densidad vehicular presente en cada microgrupo con la densidad histórica correspondiente. Aquellos con densidad superior a la media histórica se categorizan como variabilidad positiva, indicando posible congestionamiento, mientras que los de densidad inferior se clasifican como variabilidad negativa.

Agrupamiento por proximidad: Se identifican conjuntos de microgrupos cercanos dentro de un radio épsilon, formando grupos que representan zonas con patrones similares de flujo vehicular. Este proceso iterativo se detiene cuando todos los microgrupos han sido asignados a un grupo.

Sub-módulo complementario 1: Preprocesamiento de áreas de tráfico históricas

Para optimizar el análisis de la densidad vehicular, el método incorpora un sub-módulo de preprocesamiento que almacena información histórica sobre la distribución del tráfico. Estos datos pueden provenir de sistemas de información geográfica (SIG) o de registros alternativos.

La información preprocesada permite reducir la carga computacional, evitando la consulta constante a bases de datos externas. Se recomienda su actualización periódica, asegurando la vigencia de los datos utilizados para la comparación con las densidades actuales.

Sub-módulo complementario 2: Visualizador de resultados

El visualizador de resultados es una herramienta clave para la interpretación de los análisis realizados. Genera dos tipos de mapas:

Mapa interactivo por instantánea: Presenta la distribución de los microgrupos en cada momento de análisis, destacando las zonas con variabilidad positiva en rojo y las de variabilidad negativa en verde.

Mapa de distribución acumulada de variabilidades: Utiliza un sistema de reticulado para identificar celdas con alta recurrencia de congestionamiento. Cada celda recibe un índice de distribución acumulada, reflejando la frecuencia con la que se detectaron microgrupos con densidad superior a la histórica.

Novedad del método y aportes a la investigación

El método propuesto representa una contribución significativa en el ámbito del análisis del flujo vehicular, combinando estrategias de procesamiento en tiempo real con un análisis basado en datos históricos. La principal innovación radica en la integración de los dos niveles de agrupamiento, lo que permite detectar patrones de tráfico con un alto grado de precisión y adaptabilidad.

Asimismo, el uso del mecanismo de olvido porcentual garantiza que el análisis se enfoque en datos actualizados, evitando la distorsión de los resultados por información obsoleta. La combinación de estos elementos permite no solo detectar congestionamientos en tiempo real, sino también analizar tendencias a lo largo del tiempo, facilitando la toma de decisiones para la planificación urbana y la gestión del tráfico.

Resultados Obtenidos

El presente análisis se llevó a cabo en tres ciudades representativas: San Francisco, Roma y Guayaquil, utilizando datos de trayectorias vehiculares obtenidas mediante GPS. Se analizaron tres aspectos fundamentales: desplazamiento, densidad y variabilidad, con el objetivo de comprender la dinámica del tráfico en entornos urbanos con características diversas.

Para validar la efectividad del método propuesto, se implementaron métricas de evaluación en dos niveles de agrupamiento. En el Nivel 1, se utilizó el Coeficiente de Silhouette, mientras que en el Nivel 2 se aplicó el Índice de Validación de Agrupamiento Basado en Densidad (DBCV). Ambas métricas permitieron verificar la calidad y estabilidad de los grupos generados en cada ciudad, asegurando una correcta segmentación del tráfico urbano.

Análisis de los aspectos estudiados

1. Desplazamiento de los grupos

El análisis del desplazamiento del centro de masa de los grupos vehiculares permitió identificar patrones clave en la movilidad urbana. El método propuesto ajustó los desplazamientos en función de las densidades emergentes, reflejando dinámicas de congestión y dispersión del tráfico a lo largo del tiempo.

San Francisco: Se observó el mayor desplazamiento medio del centro de masa, con una alta desviación debido a la presencia de valores extremos. Esto sugiere que los grupos vehiculares tienden a reorganizarse constantemente, posiblemente por la distribución de la infraestructura vial y la presencia de vías de alta velocidad.

Guayaquil: Presentó un patrón de desplazamiento intermedio, similar al de Roma, pero con mayor variabilidad. En algunas zonas, el centro de masa se mantuvo estable, mientras que en otras hubo desplazamientos bruscos, lo que sugiere una variabilidad en la movilidad vehicular influenciada por la estructura urbana y el flujo de transporte público.

Roma: Mostró los menores desplazamientos y una distribución más estable del tráfico, con el centro de masa tendiendo a mantenerse en zonas específicas. Esto puede estar relacionado con la planificación urbana y las restricciones vehiculares en ciertas áreas.

Estos resultados reflejan la manera en que los grupos de vehículos se ajustan a la densidad del tráfico en cada ciudad, permitiendo la detección de áreas con mayor movilidad o congestión.

2. Densidad de los grupos

El estudio de la densidad vehicular permitió identificar zonas críticas de concentración de tráfico, fundamentales para evaluar la capacidad y eficiencia de la infraestructura vial.

San Francisco: Presentó la mayor dispersión en las densidades vehiculares, con valores que oscilaban entre mínimos cercanos a cero y picos extremadamente altos. Esto indica una distribución heterogénea del tráfico, donde algunas zonas experimentan alta congestión, mientras que otras permanecen con baja densidad.

Roma: Exhibió una distribución de densidad más homogénea y estable, con variaciones menores en la concentración vehicular. La planificación de la ciudad, que favorece áreas peatonales y restringe el tráfico en ciertos sectores, puede haber influido en estos resultados.

Guayaquil: Se ubicó en un punto intermedio, con una distribución de densidad menos extrema que San Francisco, pero más variable que Roma. Se detectaron zonas de alta concentración de tráfico, probablemente debido a la intersección de distintos medios de transporte (taxis, buses, motocicletas), lo que genera fluctuaciones significativas en la densidad.

Estos hallazgos evidencian la necesidad de adaptar las estrategias de gestión del tráfico a las particularidades de cada ciudad, considerando tanto la estructura vial como la variabilidad en la concentración vehicular.

3. Variabilidad de los grupos

La variabilidad vehicular analiza los cambios en la densidad a lo largo del tiempo, diferenciando entre tres tipos principales de variabilidad:

Variabilidad Positiva: Los microgrupos superan las densidades históricas, indicando un incremento significativo del tráfico.

Variabilidad Negativa: Se observa una disminución en la densidad vehicular, sugiriendo menor afluencia en ciertas áreas.

Variabilidad Mixta: Los microgrupos alternan entre períodos de alta y baja densidad, reflejando fluctuaciones complejas en la movilidad.

En San Francisco, la variabilidad positiva fue más frecuente en zonas comerciales y áreas turísticas, donde la afluencia vehicular cambia dependiendo de la hora del día y los eventos en la ciudad.

En Roma, la variabilidad fue generalmente más baja, con una estabilidad en la densidad que sugiere un flujo vehicular más constante.

En Guayaquil, se observaron altos niveles de variabilidad, con grupos vehiculares fluctuando rápidamente entre valores positivos y negativos, lo que indica una movilidad menos predecible y posiblemente afectada por eventos locales o condiciones de tráfico específicas.

El análisis de variabilidad proporciona información clave para anticipar zonas propensas a congestión y optimizar las estrategias de planificación urbana.

Validación y evaluación del agrupamiento

Para garantizar la calidad de los grupos generados, se implementaron dos métricas clave de evaluación en dos niveles de agrupamiento:

Nivel 1 - Coeficiente de Silhouette

El Coeficiente de Silhouette (SC) mide la cohesión y separación de los grupos formados, evaluando qué tan bien asignados están los elementos dentro de cada grupo. Su rango varía entre -1 y 1, donde valores cercanos a 1 indican agrupaciones bien definidas, mientras que valores cercanos a 0 o negativos sugieren asignaciones erróneas o grupos solapados.

San Francisco y Roma presentaron valores positivos y consistentes, reflejando una adecuada formación de microgrupos.

Guayaquil, aunque mostró agrupamientos razonables, presentó mayor dispersión en los resultados, lo que indica una mayor variabilidad en la calidad de los grupos formados.

Nivel 2 - Índice de Validación de Agrupamiento Basado en Densidad (DBCV)

El Índice DBCV evalúa la calidad del agrupamiento considerando dos factores principales:

Densidad intra-cluster (DSC): Mide qué tan compactos están los puntos dentro de cada grupo.

Separación inter-cluster (DSPC): Evalúa la distancia entre los grupos para determinar si están bien diferenciados.

Los valores del DBCV oscilan entre -1 y 1, donde valores altos indican una buena segmentación de los datos, mientras que valores cercanos a 0 o negativos reflejan superposiciones y baja calidad en la segmentación.

San Francisco y Roma obtuvieron valores elevados, lo que confirma una segmentación efectiva de los grupos vehiculares.

Guayaquil presentó un valor menor pero aún dentro de un rango aceptable, reflejando que los grupos formados eran adecuados pero con mayor dispersión en la estructura de densidad.

Análisis del rendimiento y tiempos de ejecución

El método fue evaluado en intervalos de 60 segundos, analizando los tiempos de procesamiento en las tres ciudades:

San Francisco: Mayor tiempo de ejecución, debido a la alta cantidad de trayectorias procesadas.

Guayaquil y Roma: Tiempos de ejecución más reducidos, con Roma mostrando la ejecución más eficiente.

La ejecución en paralelo optimizó el rendimiento, distribuyendo la carga de procesamiento de manera efectiva.

Estos resultados resaltan la eficiencia del método propuesto en el análisis del tráfico urbano, adaptándose a distintos entornos urbanos y escalando adecuadamente según la cantidad de datos procesados.

Posibles líneas de investigación futuras.

Como trabajos futuros, se plantea una serie de mejoras y ampliaciones para el método propuesto. En primer lugar, se busca aumentar su adaptabilidad en entornos urbanos complejos, enfocándose en la optimización para intersecciones y la diversidad de carreteras, así como abordar la interacción entre vehículos y peatones.

Además, se proponen casos de estudio en áreas urbanas extendidas, considerando condiciones climáticas severas y escenarios complejos, para evaluar la escalabilidad y robustez del método en contextos más amplios. Para fortalecer la comprensión del problema de congestión, se plantea explorar la influencia de la topología de la red vial en la dinámica del tráfico, analizando la interconexión entre diferentes nodos y su impacto en la congestión.

Otra línea de trabajo futuro implica la inclusión de otros modos de transporte, como vehículos privados, bicicletas y peatones, con análisis específicos por tipo de vehículo para comprender mejor las variaciones en los patrones de conducción. Asimismo, se considera esencial integrar datos multimodales, como transporte público, puntos de interés y semáforos, para obtener una visión más completa de la movilidad urbana.

En el ámbito de la inteligencia artificial, se propone implementar modelos predictivos respaldados por datos históricos y en tiempo real para anticipar y prevenir patrones de congestión. Se plantea un modelo de clasificación para la identificación de áreas congestionadas y no congestionadas, considerando variables adicionales como velocidad del flujo vehicular y capacidad de la carretera.

Otra mejora que se plantea añadir es la inclusión de una métrica para la identificar sectores con congestión vehicular por medio de un indicador que considere las características de las carreteras como la saturación de estas y pueda procesar un conjunto de datos completo de trayectorias vehiculares.

Finalmente, se sugiere validar el nuevo método en otros contextos espaciales y temporales para verificar su aplicabilidad y generabilidad, lo que permitiría mejorar la gestión del tráfico en diversas condiciones y entornos urbanos.

Estudio de Viabilidad de SYCL como Modelo de Programación Unificado para Sistemas Heterogéneos Basados en GPUs en Bioinformática.

Fecha de exposición: 21/03/2024. Institución: Universidad Nacional de La Plata.

Autor: Manuel Costanzo. mcostanzo@lidi.info.unlp.edu.ar. III-LIDI, Facultad de Informática, UNLP - CIC. La Plata, Argentina. 0000-0002-6937-3943
Director: Enzo Rucci. erucci@lidi.info.unlp.edu.ar. III-LIDI, Facultad de Informática, UNLP - CIC. La Plata, Argentina. 0000-0001-6736-7358
Director: Carlos García Sanchez. garsanca@dacya.ucm.es. Dpto. Arquitectura de Computadores y Automática, Universidad Complutense de Madrid. Madrid, España. 0000-0002-3470-1097
Co-Director: Marcelo Naiouf. mnaiouf@lidi.info.unlp.edu.ar. III-LIDI, Facultad de Informática, UNLP - CIC. La Plata, Argentina. 0000-0001-9127-3212

Resumen

Por un lado, la computación de alto rendimiento (HPC) requiere de modelos de programación que aprovechen el paralelismo masivo de arquitecturas heterogéneas modernas como las plataformas CPU-GPU. Sin embargo, modelos de programación como CUDA y OpenCL presentan limitaciones en portabilidad y productividad. SYCL surge como alternativa prometedora al unificar la programación en C++ y abstraer las particularidades del hardware. Por otro lado, la bioinformática y la biología computacional representan dos campos que han estado explotando las GPUs durante más de dos décadas, y muchas de sus implementaciones se basan en CUDA, lo que impone limitaciones significativas en cuanto a la portabilidad en una amplia gama de arquitecturas heterogéneas. Es por lo que esta tesis doctoral propone evaluar la viabilidad de SYCL como modelo de programación paralelo unificado, portable y eficiente para sistemas heterogéneos con GPUs, específicamente en bioinformática. Considerando que el alineamiento de secuencias biológicas representa una operación fundamental con amplias aplicaciones en diversas áreas de la biología y la medicina, se seleccionó la suite SW# como caso de estudio por su relevancia y por estar desarrollado en CUDA. Mediante la herramienta SYCLomatic se migró completamente el código de SW# de CUDA a SYCL. Este proceso involucró la ejecución de la herramienta, modificación del código generado, corrección de errores, verificación funcional, optimizaciones y estandarización SYCL. Posteriormente, se realizaron múltiples experimentos en un amplio conjunto de GPUs y CPUs de diferentes tipos y fabricantes para evaluar la portabilidad de rendimiento del código migrado en contextos individuales, tanto de GPU como de CPU, multi-GPU y CPU+GPU, en un entorno híbrido. Los resultados de esta tesis muestran, en primer lugar, que la herramienta SYCLomatic resulta efectiva y útil para la migración automática, aunque no pueda considerarse una solución final. En segundo lugar, las pruebas realizadas revelan que SYCL presenta una eficiencia comparable a la de CUDA en las GPUs NVIDIA y ofrece una amplia portabilidad funcional con una aceptable portabilidad de rendimiento a GPUs y CPUs de otros fabricantes. En conclusión, SYCL se posiciona como una alternativa viable como modelo unificado de programación heterogénea, portable y eficiente, estableciendo un precedente importante para futuros desarrollos en el área.

1. Introducción

1.1. Motivación

La computación de alto rendimiento (HPC) ha evolucionado considerablemente en las últimas décadas, transformándose en una solución clave para abordar problemas que requieren alta demanda

computacional como en la ciencia, la ingeniería y la industria [1]. Los avances en hardware paralelo, modelos de programación y algoritmos han permitido un crecimiento exponencial en la capacidad de procesamiento. Este cambio de paradigma ha dado lugar a arquitecturas con miles de núcleos (como *CUDA-cores* en GPUs de NVIDIA) que colaboran para potenciar el rendimiento, lo que representa un cambio de paradigma en comparación con la computación secuencial tradicional.

Este avance significativo dio lugar a la aparición y aceptación de un modelo de computación heterogénea, en el que los sistemas integran procesadores multicore de uso genérico (CPU) con aceleradores especializados como GPUs, *Field Programmable Gate Arrays* (FPGAs), T*ensor Processing Units* (TPUs), entre otros [2]; todos con sus propias aptitudes y propiedades específicas. Sin embargo, beneficiarse plenamente de este nuevo paradigma requiere más que solo avances arquitectónicos. Es esencial desarrollar software que pueda operar estas modernas y sofisticadas arquitecturas heterogéneas. En este sentido, aunque existen diversos modelos de programación, CUDA y OpenCL se destacan por su adopción generalizada, pero presentan algunas limitaciones. En particular, CUDA está restringido a las GPUs de NVIDIA, mientras que OpenCL ofrece una mayor portabilidad, pero es más complejo y menos eficiente. Alternativamente, OpenMP y OpenACC brindan opciones para la programación en paralelo, aunque con limitaciones en flexibilidad y adopción.

Debido a estas restricciones pendientes, la comunidad HPC empezó en los últimos años a indagar en la aparición y aceptación de nuevos modelos y soluciones de programación paralela que sean al mismo tiempo más sencillos, portables y eficientes. En este escenario, SYCL [3] surge recientemente como una opción especialmente prometedora para convertirse en un estándar unificado para la programación heterogénea paralela. SYCL es un modelo de programación de alto nivel basado en C++ estándar, que da la posibilidad a los programadores de crear código portable, que pueda ejecutarse de manera eficaz en arquitecturas heterogéneas que combinan CPUs con diversos tipos de aceleradores.

En la actualidad, existe un gran número de aplicaciones HPC que se desarrollaron en lenguajes y modelos más antiguos como CUDA [4] y OpenCL, por lo cual se precisa el desarrollo de técnicas efectivas para la migración de estos códigos tradicionales a SYCL. Para ello, existen herramientas como SYCLomatic [5], que están empezando a jugar un papel central en este proceso, al simplificar la automatización de la conversión de código de CUDA a SYCL.

Entre las áreas que potencialmente podrían aprovechar los beneficios de SYCL se encuentra la Bioinformática, ya que ha estado explotando el uso de aceleradores como las GPUs por más de dos décadas [6]. Son múltiples los desafíos biológicos que requieren de un alto procesamiento computacional y es por lo que existen numerosas aplicaciones de GPU para enfrentarlos, como el alineamiento de secuencias, la dinámica molecular, el acoplamiento molecular, la predicción y búsqueda de estructuras moleculares, el análisis filogenético, entre otros ejemplos. En particular, el alineamiento de secuencias biológicas (ASB) desempeña un papel fundamental en la investigación biológica, ya que se utiliza en la mayoría de sus disciplinas. Es por esto que se han desarrollado numerosas aplicaciones bioinformáticas basadas CUDA [7, 8, 9, 10, 11, 12, 13, 14, 15, 16], no obstante, presentan una clara limitación de portabilidad a otras arquitecturas. Es por ello que la adopción de SYCL en este campo puede simplificar la transición hacia un modelo más portable y

eficiente, ayudando tanto a las aplicaciones establecidas como a nuevas herramientas para que se beneficien de las ventajas de la computación heterogénea.

1.2 Objetivos, metodologías, alcances y limitaciones

El objetivo general de esta tesis consiste en evaluar la viabilidad de SYCL como un modelo de programación heterogénea unificado, portable y eficiente para el diseño y desarrollo de aplicaciones con alta demanda computacional en sistemas heterogéneos basados en GPUs, específicamente en el ámbito de la bioinformática.

Los objetivos específicos son:

- Investigar y comparar críticamente los modelos de programación y métricas de rendimiento presentes en el contexto de la computación heterogénea, específicamente en aplicaciones bioinformáticas, con el propósito de establecer una base conceptual para esta investigación.
- Diseñar y desarrollar software que aproveche las capacidades de SYCL para sistemas heterogéneos basados en GPUs en el contexto de la bioinformática, considerando especialmente la migración de aplicaciones implementadas en CUDA.
- Medir y comparar las prestaciones del software desarrollado en distintos sistemas heterogéneos basados en GPUs, considerando portabilidad, rendimiento y productividad como parámetros de interés.

Debido a limitaciones de espacio, la metodología, los alcances y las limitaciones no se encuentran incluidas en este documento, sin embargo, están disponibles para su consulta en el texto completo [17].

1.3 Contribuciones

1. Un estudio sobre la viabilidad y eficiencia del uso de SYCL para el desarrollo de aplicaciones bioinformáticas en sistemas heterogéneos basados en GPUs, con el propósito de identificar las posibilidades y métodos existentes para la creación de código, así como las ventajas y desafíos relacionados en contextos de desarrollo real. Esto es especialmente relevante para desarrolladores en el ámbito de la bioinformática (y potencialmente de otras áreas) que estén considerando la migración de una aplicación existente a SYCL, o que contemplen iniciar un nuevo proyecto desde cero utilizando esta tecnología. El estudio ayudará a identificar las posibilidades y métodos para la creación de código, así como las ventajas y desafíos asociados en contextos de desarrollo real.
2. La migración completa de la suite SW#, mediante la utilización de herramientas automáticas de migración de código CUDA a SYCL. SW# ofrece características avanzadas que permiten computar alineamientos tanto de secuencias de ADN como de proteínas; personalizar el algoritmo utilizado según su finalidad, como Smith-Waterman (SW), Needleman-Wunsch (NW), Semi-global (HW) y de Solapamiento (OV); ajustar el esquema de puntuación (matriz de sustitución más penalizaciones por *gaps*); entre otras. Al igual que la versión original, la versión migrada de SW# es capaz de combinar la potencia de cálculo de la CPU y la GPU, aunque en este caso se realiza a través del mismo lenguaje de programación. Al permitir configurar el número de hilos de la CPU y los dispositivos de la GPU a utilizar, se proporciona flexibilidad para diferentes configuraciones de hardware. Por último, SW# puede ser usada

como una herramienta independiente tanto como una librería, lo que facilita su integración en flujos de tareas bioinformáticas más amplios. Para beneficio de la comunidad científica, la migración completa de SW# a SYCL se encuentra disponible en un repositorio web público.[1]

3. Un estudio exhaustivo sobre la portabilidad funcional y de rendimiento de SYCL en el ámbito de la bioinformática. Este análisis tiene en cuenta diferentes variantes de la aplicación ASB, distintas implementaciones de SYCL y su comportamiento en sistemas basados en CPU y GPU de diferentes fabricantes, así como la combinación de ambos en un entorno híbrido. Es importante destacar que no hay estudios previos que hayan utilizado un conjunto de plataformas tan amplio y diverso en la literatura existente. Este estudio es especialmente útil para investigadores y desarrolladores en el campo de la bioinformática que buscan optimizar sus aplicaciones para un procesamiento más eficiente. Al proporcionar un análisis exhaustivo sobre las variaciones en el rendimiento de SYCL en distintas configuraciones de hardware, este estudio contribuye a determinar si SYCL representa una alternativa viable como modelo de programación unificado para la computación heterogénea. Además, este análisis también beneficia a la comunidad académica, proporcionando un recurso educativo para aquellos interesados en aprender sobre la portabilidad del rendimiento en diferentes entornos de cómputo y su aplicación práctica en la bioinformática.

2. Marco Referencial

Esta sección presenta los conceptos y tecnologías relevantes para este trabajo.

2.1 SYCL

SYCL es un modelo de programación abstracto y multiplataforma en C++ diseñado específicamente para la computación heterogénea. Su desarrollo surge debido a la necesidad de ofrecer una alternativa que combine la portabilidad y eficiencia de APIs paralelas como OpenCL con la facilidad de uso y flexibilidad del C++ moderno, lo que permite a los desarrolladores escribir código estándar en este lenguaje utilizando técnicas habituales de alto nivel, al tiempo que acceden a un amplio rango de capacidades de implementaciones subyacentes como OpenCL [3].

A diferencia de otras soluciones como CUDA, que requieren un lenguaje de programación específico o un entorno de ejecución particular, SYCL se integra de manera más orgánica en el flujo de trabajo de desarrollo en C++, lo que facilita su adopción en proyectos existentes. Por otro lado, en contraste con soluciones para la computación heterogénea como OpenACC, que se centra en la simplicidad mediante directivas de compilación, SYCL ofrece una mayor flexibilidad y control detallado de los dispositivos, permitiendo a los desarrolladores aprovechar mejor las capacidades del hardware y personalizar el cómputo paralelo según las necesidades específicas de sus aplicaciones.

Existen diversas implementaciones de SYCL, cada una optimizando su aplicabilidad para diferentes plataformas: DPC++ de Intel, parte del ecosistema de programación unificado oneAPI, ofrece soporte para CPUs, GPUs y FPGAs de Intel con herramientas avanzadas de desarrollo. ComputeCpp, ahora integrada en oneAPI, brinda una solución comercial compatible con múltiples plataformas. AdaptiveCpp, permite la programación en CPU y GPU con un enfoque de compilación único que facilita la portabilidad entre plataformas y proveedores. TriSYCL, de Xilinx, es una implementación

[1] https://github.com/ManuelCostanzo/swsharp_sycl

de código abierto centrada en FPGAs y en la experimentación con extensiones de SYCL. Estas opciones resaltan la flexibilidad de SYCL para diferentes entornos y dispositivos.

2.2 Portabilidad de rendimiento

La evaluación de la portabilidad de rendimiento es fundamental para asegurar que una aplicación aproveche eficientemente los recursos en múltiples plataformas y se adapte a los avances tecnológicos en HPC. Se compone de dos aspectos: la portabilidad funcional, que garantiza el correcto funcionamiento de una aplicación en distintos sistemas operativos y arquitecturas de hardware, y la portabilidad de rendimiento, que se enfoca en mantener un rendimiento óptimo en diversas configuraciones. Esto implica medir la eficiencia de una aplicación, comparándola con el mejor rendimiento conocido para esa plataforma, y la eficiencia arquitectónica, que mide cómo se aprovechan los recursos en relación con el rendimiento máximo teórico o práctico. Una evaluación adecuada de estas métricas ayuda a identificar cuellos de botella y a optimizar el código para un mejor rendimiento en cada entorno.

Penycook *et al.* [18] presentan una métrica para medir la portabilidad de rendimiento, pero es reformulada por Marowka [19] luego de haber identificado limitaciones y sesgos en la misma. Este enfoque considera tanto la eficiencia de la aplicación como la eficiencia arquitectónica, permitiendo comparaciones más justas entre plataformas.

2.3 Alineamiento de secuencias biológicas (ASB)

El ASB es una herramienta esencial en bioinformática y biología molecular, utilizada para entender relaciones evolutivas y funcionales entre secuencias de proteínas y ácidos nucleicos. Inicialmente, se desarrolló para identificar homologías entre secuencias, facilitando el análisis filogenético, la detección de enfermedades genéticas y la predicción de secuencias ancestrales. Para mejorar la precisión del ASB, se introducen matrices de puntuación y sustitución, que asignan valores a coincidencias, discrepancias y *gaps*, optimizando la identificación de relaciones evolutivas y funcionales.

Los algoritmos para ASB se clasifican en globales, que comparan secuencias en su totalidad, y locales, que identifican regiones con alta similitud. Los globales son útiles para secuencias de longitud similar con relaciones evolutivas cercanas, mientras que los locales encuentran regiones conservadas sin requerir una relación cercana general. Los métodos se implementan utilizando programación dinámica o heurística, siendo cada uno más adecuado según las características y necesidades del alineamiento.

Dos de los algoritmos basados en programación dinámica más conocidos son Needleman-Wunsch (NW) para alineamiento global y Smith-Waterman (SW) para el local. También existen alternativas como el alineamiento semi-global (HW) y el de solapamiento (OV), que siguen las etapas de inicialización, relleno de la matriz y retroceso para encontrar el mejor alineamiento.

Para evaluar el rendimiento en el contexto de ASB, se usa una métrica específica denominada GCUPS. Por definición, un CUPS representa la cantidad de tiempo necesario para computar una celda completa de la matriz de similitud. Considerando una secuencia de consulta Q y una base de datos D, el rendimiento en GCUPS se calcula mediante la fórmula: $GCUPS = \frac{|Q| \times |D|}{t \times 10^9}$ donde $|Q|$ denota el

número total de símbolos en la secuencia de consulta, $|D|$ representa el número total de símbolos en la base de datos, y t es el tiempo total de ejecución expresado en segundos.

3. Caso de Estudio y Migración a SYCL

Esta sección se presenta el caso de estudio seleccionado y el análisis de su migración a SYCL.

3.1 Selección de software para ASB

Se seleccionó la suite bioinformática SW# como caso de estudio debido a su capacidad para computar ASB. SW# permite configuraciones avanzadas para alinear pares de secuencias de proteínas y ADN, ofreciendo soporte para diversos tipos de algoritmos, incluyendo SW, NW, HW y OV. Además, integra técnicas de optimización como la distribución dinámica de la carga de trabajo y el acceso eficiente a la memoria de la GPU, combinando el procesamiento concurrente de CPU y GPU para un rendimiento óptimo. SW# puede utilizarse como una herramienta independiente o como una librería que facilita la integración en flujos de trabajo más amplios. Al estar originalmente desarrollado en CUDA, representa un candidato ideal para estudiar la migración a SYCL y así evaluar la portabilidad y rendimiento en diversas plataformas.

3.2 Herramientas y frameworks

oneAPI es una iniciativa de Intel que tiene como objetivo simplificar el desarrollo de aplicaciones para diversas arquitecturas de hardware, como CPUs, GPUs y FPGAs. Su ecosistema incluye herramientas de programación directa y basadas en API, ofreciendo una interfaz unificada para trabajar con diferentes tipos de hardware. La portabilidad, la heterogeneidad y el rendimiento son características fundamentales de oneAPI, que proporciona un amplio conjunto de herramientas y librerías para facilitar el desarrollo de aplicaciones heterogéneas [20].

La herramienta de migración SYCLomatic, creada por Intel como parte de oneAPI, automatiza la conversión de código CUDA a SYCL. Esta herramienta resulta útil para migrar código existente y aprovechar las ventajas de SYCL sin necesidad de reescribir por completo las aplicaciones. Es importante tener en cuenta que SYCLomatic es una herramienta que asiste en la migración de código CUDA a SYCL, pero no garantiza una conversión perfecta en todos los casos. En particular, suele migrar entre un 80% a un 90% del código, dependiendo de la complejidad del mismo, por lo cual es posible que aún sea necesario realizar ajustes manuales en el código resultante para asegurar un comportamiento correcto y un rendimiento óptimo en el nuevo entorno SYCL.

3.3 Proceso de migración

El proceso de migración en esta tesis se compone de 6 etapas, de las cuales las últimas 2 son opcionales. En primer lugar, se emplea la herramienta SYCLomatic para generar el código correspondiente. A continuación, se procede a corregir los errores y advertencias con el fin de obtener una versión ejecutable. Posteriormente, se solucionan los errores de tiempo de ejecución para lograr una versión funcional. Luego, se verifica la precisión de los resultados obtenidos y, de manera opcional, se puede estandarizar y optimizar el código resultante.

La Figura 1 resume las advertencias generadas por SYCLomatic, agrupadas en 4 áreas: manejo de errores (DPCT1003), funcionalidades no soportadas (DPCT1005, DPCT1084 y DPCT1059), recomendaciones (DPCT1049) y optimizaciones (DPCT1065).

3.3.1 Verificación funcional

Tras finalizar el proceso de migración, se realizaron diferentes pruebas, tanto para secuencias de proteínas como de ADN, utilizando distintos algoritmos de alineamiento y esquemas de puntuación. Finalmente, se verificó que tanto CUDA como SYCL produjeran los mismos resultados.

3.3.2 Modernización de código y optimizaciones

El código original SW# fue desarrollado exclusivamente para GPUs de NVIDIA con configuraciones estáticas como las dimensiones de los bloques para los kernels, limitando su aprovechamiento en GPUs actuales de mayor capacidad e impidiendo su ejecución en dispositivos con diferentes requisitos de grupo de trabajo (WG, por sus siglas en inglés), como GPUs de otros fabricantes o CPUs. Para solventar esto, se reemplazó la configuración estática del tamaño del WG por una configuración dinámica que considera las longitudes de secuencia y el valor máximo permitido por el dispositivo correspondiente, extendiéndose así el soporte del código migrado a diversas arquitecturas, al tiempo que también se aplicó la misma mejora al código CUDA original para una evaluación de rendimiento justa.

3.3.3 Estandarización a SYCL (opcional)

Si bien SYCLomatic produce código DPC++, en algunos casos genera dependencias explícitas a funciones internas de oneAPI. Para lograr un código SYCL completamente portable, se realizaron ajustes manuales donde se elimnaron estas dependencias con el objetivo de tener una versión SYCL estandar y así poder utilizar cualquier implementación SYCL disponible, aumentando la portabilidad.

3.3.4 Evaluación del esfuerzo de programación

El esfuerzo de programación en este proceso de migración de código está directamente relacionado con la eficiencia de SYCLomatic. Para esto, llevó a cabo una evaluación de la traducción automática mediante la medición de las SLOC de las versiones de CUDA y SYCL. La Tabla 1 muestra las SLOC en las diferentes etapas del proceso.

El análisis del proceso de migración con SYCLomatic muestra una eficiencia destacada en la traducción automática del código de CUDA a SYCL, logrando migrar un 95% del código, confirmando lo promocionado por Intel en cuanto a la efectividad de la herramienta. No obstante, esta eficiencia se ve atenuada por la necesidad de intervenciones manuales significativas. A pesar de este requisito adicional, la herramienta demuestra ser una solución valiosa en el ámbito de la migración de código, ya que facilita en gran medida la transición entre plataformas de cómputo.

3.4 Trabajos relacionados y discusión

Se estudiaron y compararon 13 trabajos relacionados de migración de código existente CUDA a SYCL. Como conclusión general, la herramienta SYCLomatic facilita la migración de código entre ambos modelos, automatizando la conversión de patrones comunes y reduciendo errores humanos, lo que ahorra tiempo y esfuerzo. Sin embargo, no es una solución completa, ya que requiere intervención manual en funciones o segmentos sin equivalentes directos en SYCL. Aunque no es indispensable, SYCLomatic simplifica la migración y los desarrolladores pueden optar por realizar una conversión manual para aplicar optimizaciones particulares o lograr un código más limpio, pero el esfuerzo de programación será considerablemente mayor.

4. Resultados Experimentales

Esta sección se centra en la presentación de los resultados obtenidos en cuanto a la portabilidad funcional y de rendimiento de SW# en las diferentes plataformas.

4.1 Diseño experimental

La evaluación se realizó en un conjunto de 8 dGPUs (6 de NVIDIA, 1 de AMD, 1 de Intel), 3 iGPUs (2 de Intel, 1 de AMD) y 8 CPUs (7 de Intel, 1 de AMD). Las características completas de cada plataforma se encuentran en la tesis [17].

SW# se ejecutó tanto en CUDA y SYCL, utilizando las implementaciones de oneAPI y AdaptiveCpp. Para la evaluación en proteínas se usaron las bases de datos Swiss-Prot y Env. NR (7 veces más grande que la anterior), mientras que para ADN se seleccionaron 5 pares de secuencias de diferentes tamaños.

Las pruebas se realizaron en diversos escenarios de SW#, que incluyeron diferentes configuraciones de matrices de puntuación y de sustitución, penalizaciones y distintos algoritmos de alineamiento. Las pruebas se ejecutaron 20 veces y se deshabilitó la cooperación de la CPU para reducir el impacto de la misma. Además, se seleccionó el tamaño de WG óptimo, logrando una comparación precisa y minimizando la variabilidad.

4.2 Resultados de rendimiento y portabilidad

En esta sección se presenta el análisis de los resultados obtenidos.

4.2.1 Rendimiento y funcionalidad

A continuación, se presentan las diferentes comparaciones de rendimientos entre CUDA y SYCL para la suite completa de SW#, bajo los siguientes escenarios:

Tamaño del WG: la Figura 2a muestra que ambos modelos son sensibles al tamaño del WG, siendo la configuración 1024 y dinámica las que mejores rendimientos logran. Como se puede apreciar, las GPUs más recientes demostraron alcanzar un mayor rendimiento de GCUPS con esta configuración. Los siguientes experimentos se ejecutan bajo la configuración dinámica.

Bases de datos y longitudes de secuencias: por un lado, la Figura 2b revela una ligera pérdida de rendimiento al utilizar bases de datos más grandes, aunque esta disminuye o incluso se revierte con GPUs más potentes. CUDA mostró un rendimiento superior en algunos casos, pero las diferencias entre CUDA y SYCL generalmente no superan el 2%, lo que indica un rendimiento comparable entre ambos modelos. Por otro lado, la Figura 2c muestra que secuencias de consulta más largas tienden a mejorar el rendimiento, especialmente en las GPUs más potentes como la RTX 3090, que requiere un volumen de trabajo suficiente para aprovechar al máximo su potencia.

Algoritmos de alineamiento y matrices de puntuación: la diferencia de rendimiento entre los distintos algoritmos de alineamiento y matrices de puntuación fue de alrededor del 2% en promedio (Figuras 2d y 2e, lo que indica que estos parámetros no afectan significativamente el rendimiento del código migrado a SYCL. Esto confirma la funcionalidad y el rendimiento comparable de SYCL con CUDA en este experimento.

Alineamiento de secuencias de ADN: la Figura 2f demuestra que las secuencias de ADN más largas no siempre llevan a un mayor rendimiento en GCUPS, lo cual podría deberse al grado de similitud

entre secuencias. SYCL supera a CUDA por un 10% en promedio en la GPU RTX 2070, mientras que en la V100 esta ventaja disminuye a un 7%. Se realizó perfilado más detallado en la GPU RTX 2070, que indica que SYCL supera a CUDA en la gestión de memoria y productividad computacional. Sin embargo, en la RTX 3090, las métricas son casi iguales, lo que podría deberse a diferencias en la microarquitectura entre estas GPUs.

4.2.2 Modelo para portabilidad de rendimiento

Para evaluar la portabilidad de rendimiento, se estima el rendimiento teórico máximo de GPUs y CPUs en base al modelo desarrollado por Lan *et al* [21]. Para ello, se consideran características de hardware y algoritmos, calculando la capacidad de cómputo a partir de la frecuencia de reloj, el rendimiento por ciclo y la cantidad de carriles SIMD. Luego, se mide el número de instrucciones para actualizar una celda en la matriz de similitud, permitiendo modelar el rendimiento máximo teórico de un dispositivo en alineamiento de secuencias. El modelo se adaptó para SW# y se extendió a GPUs de AMD e Intel, tanto discretas como integradas, además de CPUs de ambos fabricantes.

4.2.3 Instrucciones *core* de SW# y características arquitectónicas de GPUs y CPUs

De acuerdo con mediciones realizadas, el 99% del tiempo de ejecución de SW# es destinado al cómputo de las matrices de similitud. Para dicho cómputo utiliza enteros de 32 bits y realiza 12 instrucciones por actualización de celda. Por otra parte, se llevó a cabo un minucioso estudio de las arquitecturas empleadas para determinar su capacidad de cómputo a partir de la frecuencia de reloj, la cantidad de carriles SIMD y el rendimiento asociado a cada uno. Los detalles completos se encuentran en la tesis [17].

4.2.4 Portabilidad en GPU individual

En la comparación inicial entre CUDA y SYCL en las GPUs de NVIDIA, ambos modelos demostraron rendimientos prácticamente idénticos en GCUPS. SYCL superó a CUDA en la Tesla V100 (3.4%), mientras que CUDA fue mejor en la GTX 980 (4.6%), lo que representa que ambas plataformas ofrecen un rendimiento comparable en NVIDIA. La comparación detallada de GPUs de NVIDIA, AMD e Intel se presenta en la Tabla 2, donde evidencia que las GPUs más potentes alcanzan mayores valores en GCUPS, con eficiencias arquitectónicas entre el 37% y el 52%, siendo la más destacada la RTX 2070. En las GPUs de AMD e Intel, SYCL alcanzó una eficiencia arquitectónica del 51.7% en la dGPU de AMD, mientras que las iGPUs superaron esta marca. Sin embargo, la eficiencia decrece en la dGPU Arc A770 de Intel (23.3%) y la iGPU Vega 6 de AMD (21.3%), posiblemente debido a diferencias en la filosofía de diseño.

En cuanto a la portabilidad del rendimiento presentada en la Tabla 4, SYCL mostró tasas similares a la de CUDA en las GPUs de NVIDIA, con valores de 42% y 42.2%, respectivamente. Sin embargo, SYCL fue más eficiente en las iGPUs de Intel, mientras que en AMD se observó una buena portabilidad en la dGPU. En general, SYCL demostró una mayor portabilidad al lograr una alta eficiencia en GPUs de varios fabricantes. Además, obtuvo una eficiencia arquitectónica superior en tres de los cinco casos probados, destacando la capacidad de SYCL para mejorar el rendimiento en una amplia gama de GPUs.

4.2.5 Portabilidad en multi-GPU

La comparación entre CUDA y SYCL utilizando diversas combinaciones multi-GPU de NVIDIA demuestra que el rendimiento es prácticamente idéntico. CUDA superó a SYCL en una configuración de 2xGTX1080 por alrededor del 1%, mientras que SYCL fue superior en las demás combinaciones, alcanzando hasta un 5% más de GCUPS. Esto indica que SYCL no tiene una sobrecarga adicional al emplear múltiples GPUs. La Tabla 3 muestra que las tasas de eficiencia arquitectónica fueron ligeramente menores en configuraciones multi-GPU debido a la distribución de carga de trabajo sencilla por parte de SW#, que no tiene en cuenta la potencia de cada GPU ni equilibra la longitud de las secuencias de consulta. Es destacable también que SYCL demostró portabilidad funcional al ejecutar 2 GPUs de Intel de diferentes tipos (iGPU y dGPU) aunque con un rendimiento no óptimo.

4.2.6 Portabilidad en CPU

La Tabla 8 demuestra la portabilidad funcional de SYCL en una amplia gama de CPUs, incluyendo modelos de Intel y AMD, a pesar de las diferencias arquitectónicas entre GPUs y CPUs. SW#, originalmente diseñado para GPUs, fue capaz de ejecutarse en CPUs sin cambios significativos, validando la etapa 5 del proceso de migración. Además, fue compatible con arquitecturas híbridas como la Core i9-13900K. Aunque las CPUs ofrecen menos GCUPS que las GPUs por naturaleza, 7 de las 8 CPUs evaluadas superaron el 38% de eficiencia arquitectónica.

La Tabla 5 muestra la portabilidad del rendimiento en CPUs. Las CPUs Intel de segmento de escritorio tuvieron una ligera ventaja sobre las de servidor, mientras que en términos generales, Intel superó a AMD. Esta diferencia podría atribuirse a que el ecosistema oneAPI está optimizado para procesadores Intel. Cabe destacar que la portabilidad de rendimiento fue del 42.2% en 6 GPUs NVIDIA, mientras que en CPUs alcanzó el 41.39%, una diferencia menor al 1%, demostrando la capacidad de SYCL para ejecutarse eficientemente en diferentes arquitecturas.

4.2.7 Portabilidad en CPU-GPU

Se evaluó la portabilidad de SYCL en combinaciones CPU-GPU, con el objetivo principal de validar la portabilidad funcional. La Tabla 6 muestra los GCUPS y eficiencias arquitectónicas en diversas combinaciones, reflejando la escalabilidad del código en arquitecturas de NVIDIA. Incluso las combinaciones con menor rendimiento, como VEGA6 y RYZEN 3, evidencian la portabilidad funcional de SYCL en hardware con recursos limitados.

La eficiencia arquitectónica varía según las combinaciones CPU-GPU, destacando la integración efectiva entre las GPUs de NVIDIA y CPUs de Intel. Las combinaciones con GPUs de Intel o AMD muestran eficiencias arquitectónicas menores, indicando una oportunidad para mejorar la sinergia entre estas arquitecturas. La Tabla 7 revela que la combinación de CPU Intel y GPU NVIDIA tiene la mejor eficiencia, con un 27.39%, mientras que la combinación de CPU Intel y dGPU AMD tiene el menor rendimiento, con un 5.97%. Las iGPUs de AMD e Intel presentan eficiencias intermedias entre el 14.03% y el 17.80%. Cabe destacar que el rendimiento global se ve afectado al incluir CPUs, que por naturaleza son más lentas que las GPUs, lo que resalta la necesidad de optimizar la distribución de la carga de trabajo en SW#. Este enfoque debe considerar las capacidades de cada dispositivo para evitar una asignación ineficiente, permitiendo futuras optimizaciones en entornos híbridos CPU-GPU.

4.2.8 Portabilidad en implementaciones SYCL

Se seleccionó AdaptiveCpp para verificar la portabilidad entre diferentes implementaciones de SYCL. AdaptiveCpp permite compilar código SYCL de forma genérica para apuntar a cualquier arquitectura soportada o en modo específico, donde se debe indicar a qué plataformas apuntar. La Figura 9 muestra el rendimiento de ambas implementaciones con la base de datos Swiss-Prot. Se observaron pérdidas de hasta un 15% en AdaptiveCpp al usar el compilador genérico en lugar del específico. No obstante, las diferencias entre oneAPI y AdaptiveCpp fueron insignificantes, a excepción de la GPU Arc A770, donde oneAPI fue solo un 5% más óptimo.

Esto evidencia la interoperabilidad de SYCL y su capacidad para mantener un rendimiento uniforme entre implementaciones, permitiendo a los desarrolladores elegir la que mejor se adapte a sus necesidades sin comprometer el rendimiento, algo fundamental en HPC.

4.3 Trabajos relacionados y discusión

Se estudiaron 18 trabajos relacionados y, en general, los resultados de rendimiento entre las implementaciones de SYCL fueron similares, con algunas diferencias aisladas. En este estudio, SW# fue migrado con mínima intervención manual para analizar la portabilidad funcional y de rendimiento de SYCL, abarcando un conjunto de plataformas más diverso que cualquier trabajo previo. La portabilidad funcional se verificó en varias arquitecturas de GPU y CPU de NVIDIA, Intel y AMD. El análisis consideró variantes de la aplicación ASB y diversas implementaciones de SYCL en configuraciones puras e híbridas de CPU y GPU. El código migrado demostró portabilidad y rendimiento sin overhead en GPUs de NVIDIA y manteniendo eficiencia en GPUs y CPUs de otros proveedores, aunque en configuraciones multi-GPU y CPU-GPU la eficiencia disminuyó. Estos resultados muestran que SYCL permite portar aplicaciones a diversas plataformas sin reescribir el código, aumentando la productividad. A pesar de algunas limitaciones, SYCL está evolucionando rápidamente y su comunidad sigue creciendo, consolidándolo como un modelo de programación unificado, portable y eficiente para sistemas heterogéneos en bioinformática.

5. Conclusiones

Esta tesis se ha planteado como objetivo general *evaluar la viabilidad de SYCL como un modelo de programación heterogénea unificado, portable y eficiente para el diseño y desarrollo de aplicaciones con alta demanda computacional en sistemas heterogéneos basados en GPUs, específicamente en el ámbito de la bioinformática*. Para esto, se investigaron críticamente modelos de programación como CUDA, OpenCL, OpenMP, OpenACC, Kokkos, RAJA y SYCL, junto con métricas clave como rendimiento, portabilidad y esfuerzo de programación. Además, se estudió el ASB por representar una operación fundamental con amplias aplicaciones en diversas áreas de la biología y la medicina. La migración completa del código CUDA de la suite SW# a SYCL se logró utilizando la herramienta SYCLomatic, que luego de un riguroso proceso de migración, resultó en una versión completamente funcional adecuada para sistemas heterogéneos.

Se llevaron a cabo múltiples experimentos para evaluar la portabilidad funcional y de rendimiento de SYCL en diferentes contextos, incluyendo diversas combinaciones de GPUs de NVIDIA, Intel y AMD, tanto discretas como integradas, así como CPUs de Intel y AMD. Se analizaron configuraciones individuales e híbridas de CPU-GPU y se evaluó AdaptiveCPP como implementación SYCL alternativa.

En conclusión, SYCL demostró una notable portabilidad funcional y de rendimiento en varias arquitecturas, alcanzando eficiencias arquitectónicas comparables en estos distintos dispositivos. Aunque las configuraciones multi-GPU y CPU-GPU mostraron una disminución en la eficiencia debido a factores externos a SYCL, el código migrado ofreció un rendimiento consistente en general. Esto establece a SYCL como un modelo de programación unificado para sistemas heterogéneos en bioinformática, que si bien presenta limitaciones específicas, la comunidad está en constante crecimiento para superar estas barreras.

De acuerdo a los resultados obtenidos y a las contribuciones realizadas, se espera que esta tesis contribuya al avance de la programación unificada para sistemas heterogéneos basados en GPUs en bioinformática, abriendo nuevas oportunidades para el desarrollo de aplicaciones eficientes y portables en este campo.

Trabajos Futuros

- Optimizar el código SYCL para alcanzar su máximo rendimiento. En particular, la suite original SW# no considera algunas optimizaciones conocidas para el alineamiento SW, como la reordenación de instrucciones para reducir la cantidad de las mismas y el uso de enteros de menor precisión para aumentar la paralelización. Además, se busca mejorar la estrategia de distribución de carga de trabajo al utilizar más de un dispositivo. Estas mejoras conducirán a tasas de eficiencia más altas.
- Ejecutar el código SYCL en otras arquitecturas FPGAs y considerar otros modelos de programación como Kokkos y RAJA, para fortalecer el estudio actual de portabilidad de rendimiento..

Publicaciones derivadas de la tesis:

1. ***"Early Experiences Migrating CUDA codes to oneAPI"***. Manuel Costanzo, Enzo Rucci, Carlos García-Sanchez y Marcelo Naiouf. *Short papers of the 9th Conference on Cloud Computing Conference, Big Data & Emerging Topics*. Págs 14-18. Mayo de 2021. https://sedici.unlp.edu.ar/handle/10915/125138
2. ***Migrating CUDA to oneAPI: A Smith-Waterman Case Study"***. Manuel Costanzo, Enzo Rucci, Carlos García-Sanchez, Marcelo Naiouf, Manuel Prieto-Matías. *In: Rojas, I., Valenzuela, O., Rojas, F., Herrera, L.J., Ortuño, F. (eds) Bioinformatics and Biomedical Engineering. IWBBIO 2022. Lecture Notes in Computer Science(), vol 13347. Springer, Cham. https://doi.org/10.1007/978-3-031-07802-6_9*
3. ***"Brief Performance Portability Analysis of a Matrix Multiplication Kernel on Multiple Vendor GPUs"***. Manuel Costanzo, Enzo Rucci, Carlos García-Sanchez y Marcelo Naiouf. *XI Jornadas de Cloud Computing, Big Data & Emerging Topics*. Págs 13-18. Junio de 2023. http://sedici.unlp.edu.ar/handle/10915/155420
4. ***"Comparing Performance and Portability between CUDA and SYCL for Protein Database Search on NVIDIA, AMD, and Intel GPUs"***. Manuel Costanzo, Enzo Rucci, Carlos García-Sanchez, Marcelo Naiouf, Manuel Prieto-Matías. 2023 *IEEE 35th International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD)*. Págs. 141-148. Junio de 2023. DOI: 10.1109/SBAC-PAD59825.2023.00023. https://www.computer.org/csdl/proceedings-article/sbac-pad/2023/054800a141/1SBGSAqfmqQ

5. ***"Assessing Opportunities of SYCL for Biological Sequence Alignment on GPU-based Systems"***. Manuel Costanzo, Enzo Rucci, Carlos García-Sanchez, Marcelo Naiouf, Manuel Prieto-Matías. *The Journal of Supercomputing*. Febrero de 2024. DOI: 10.1007/s11227-024-05907-2. https://link.springer.com/article/10.1007/s11227-024-05907-2

6. ***"Analyzing the Performance Portability of SYCL across CPUs, GPUs, and Hybrid Systems with SW Sequence Alignment".*** Manuel Costanzo, Enzo Rucci, Carlos García-Sanchez, Manuel Prieto-Matías. Future Generation Computer Systems. Under Review (R1).

Referencias

[1] H. Giefers, P. Staar, C. Bekas, and C. Hagleitner, "Analyzing the energy-efficiency of sparse matrix multiplication on heterogeneous systems: A comparative study of gpu, xeon phi and fpga," in 2016 IEEE ISPASS, 2016, pp. 46–56.
[2] E. Rucci, "Evaluación de rendimiento y eficiencia energética de sistemas heterogéneos para bioinformática," Tesis Doctoral, UNLP, 2016.
[3] Khronos SYCL working group, "Sycl 2020 specification," https://registry.khronos.org/SYCL/specs/sycl-2020/html/sycl-2020.html, 2023.
[4] Robert Dow, "GPU shipments increase year-over-year in Q3," 2021, https://www.jonpeddie.com/press-releases/gpu-shipments-increase-year-over-year-in-q3.
[5] Intel. Syclomatic: A new cuda∗-to-sycl∗code migration tool. [Online]. Available: https://www.intel.com/content/www/us/en/developer/articles/technical/syclomatic-new-cuda-to-sycl-code-migration-tool.html
[6] M. S. Nobile, P. Cazzaniga, A. Tangherloni, and D. Besozzi, "Graphics processing units in bioinformatics, computational biology and systems biology," Briefings in Bioinformatics, vol. 18, no. 5, pp. 870–885, 07 2016.
[7] E. F. O. Sandes and A. C. M. de Melo, "Cudalign: Using gpu to accelerate the comparison of megabase genomic sequences," in Proceedings of the 15th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming, 2010.
[8] E. Sandes and A. Melo, "Smith-waterman alignment of huge sequences with gpu in linear space," 06 2011, pp. 1199–1211.
[9] ——, "Retrieving smith-waterman alignments with optimizations for megabase biological sequences using gpu," Parallel and Distributed Systems, IEEE Transactions on, 05 2013.
[10] E. Sandes, G. Miranda, A. Melo, X. Martorell, and E. Ayguade, "Cudalign 3.0: Parallel biological sequence comparison in large gpu clusters," 05 2014, pp. 160–169.
[11] E. F. d. O. Sandes, G. Miranda, X. Martorell, E. Ayguade, G. Teodoro, and A. C. M. Melo, "Cudalign 4.0: Incremental speculative traceback for exact chromosome-wide alignment in gpu clusters," IEEE Transactions on Parallel and Distributed Systems, 2016.
[12] M. Korpar and M. Sikic, "SW# - GPU-enabled exact alignments on genome scale." Bioinformatics, vol. 29, no. 19, pp. 2494–2495, 2013.
[13] Y. Liu, D. L. Maskell, and B. Schmidt, "CUDASW++: optimizing Smith-Waterman sequence database searches for CUDA-enabled graphics processing units," BMC Res. Notes, vol. 2, no. 1, p. 73, May 2009.
[14] Y. Liu, B. Schmidt, and D. L. Maskell, "CUDASW++2.0: enhanced Smith-Waterman protein database search on CUDA-enabled GPUs based on SIMT and virtualized SIMD abstractions," BMC Res. Notes, vol. 3, no. 1, p. 93, Apr. 2010.
[15] Y. Liu, A. Wirawan, and B. Schmidt, "Cudasw++ 3.0: accelerating smith-waterman protein database search by coupling cpu and gpu simd instructions," BMC Bioinformatics, 2013.
[16] B. Schmidt, F. Kallenborn, A. Chacón, and C. Hundt, "Cudasw++4.0: Ultra-fast gpu-based smith-waterman protein sequence database search," 10 2023.
[17] M. Costanzo, Estudio de viabilidad de SYCL como modelo de programación unificado para sistemas heterogéneos basados en GPUs en bioinformática, 2024. [Online]. Available: https://sedici.unlp.edu.ar/handle/10915/164928
[18] S. Pennycook, J. Sewall, and V. Lee, "Implications of a metric for performance portability," Future Generation Computer Systems, vol. 92, pp. 947–958, 2019.
[19] A. Marowka, "Reformulation of the performance portability metric," Software: Practice and Experience, vol. 52, no. 1, pp. 154–171, 2022.
[20] I. Corporation, "Intel oneapi programming guide," https://community.intel.com/legacyfs/online/drupal_files/oneAPIProgrammingGuide_9.pdf, 2023.
[21] H. Lan, W. Liu, Y. Liu, and B. Schmidt, "Swhybrid: A hybrid-parallel framework for large-scale protein sequence database search," in 2017 IEEE International Parallel and Distributed Processing Symposium (IPDPS), 2017, pp. 42–51.

Tablas y Figuras

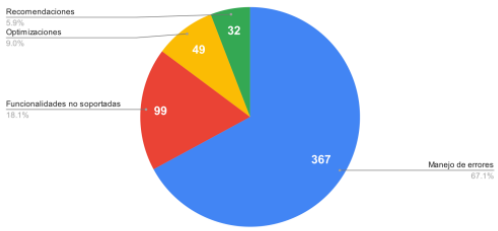


Figura 1: Distribución de las advertencias generadas por SYCLomatic.

Tabla 1: SLOC de SW# (versiones CUDA y SYCL).

CUDA[1]		SYCL[2]		
Original	No migrado	Resultado SYCLomatic	Manuales	SYCL puro
8072	408	10457	12175	9866

Tabla 4: Portabilidad del rendimiento de los códigos CUDA y SYCL en GPUs individuales.

Conjunto de plataformas (H)	$\Phi(\alpha, p, H)$ **CUDA**	**SYCL**
NVIDIA	42 %	42.2 %
AMD (discreto)	NA	51.7 %
AMD (integrado)	NA	21.3 %
AMD (todos)	NA	36.5 %
Intel (discreto)	NA	23.3 %
Intel (integrado)	NA	72.0 %
Intel (todos)	NA	55.8 %
NVIDIA ∪ AMD	NA	40.7 %
NVIDIA ∪ Intel	NA	47.2 %
Intel ∪ AMD	NA	48.1 %
NVIDIA ∪ AMD ∪ Intel	NA	44.8 %

Tabla 5: Portabilidad del rendimiento de los códigos CUDA y SYCL en CPUs individuales.

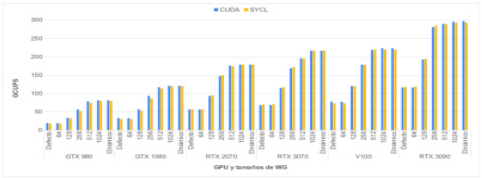
Conjunto de plataformas (H)	$\Phi(\alpha, p, H)$ **CUDA**	**SYCL**
Intel CPUs (escritorio)	NA	44.35 %
Intel CPUs (servidor)	NA	38.30 %
AMD CPU (mobile)	NA	38.82 %
Intel CPUs	NA	41.76 %
AMD CPU	NA	38.82 %
Intel ∪ AMD	NA	41.39 %

Tabla 6: GCUPS y eficiencias arquitectónicas de los códigos CPU-GPU en múltiples combinaciones.

Comb.	Plataforma: GPU U CPU	GCUPS pico	CUDA: GCUPS obt.	CUDA: Arq efi.	SYCL: GCUPS obt.	SYCL: Arq efi.
NVIDIA U Intel CPU	GTX980 + E52695	191	×	NA	73	38.1%
	GTX1080 + E52695	287	×	NA	84	29.5%
	RTX2070 + i57400	320	×	NA	49	15.3%
	RTX3070 + i5-10400F	439	×	NA	149	33.9%
	RTX3090 + Gold 6138	843	×	NA	210	24.9%
	V100 + Gold 6138	690	×	NA	156	22.6%
Intel GPU U Intel CPU	ARC770 + i9-13900K	859	×	NA	124	14.5%
	UHD770 + i9-13900K	75	×	NA	13	17.8%
	ARC770 + UHD770 + i9-13900K	894	×	NA	128	14.3%
AMD GPU U Intel CPU	RX6700 + E5-1620	560	×	NA	33	6.0%
AMD GPU U Intel CPU	VEGA6 + RYZEN 3	45.0	×	NA	6	14.0%

Tabla 7: Portabilidad del rendimiento de los códigos CUDA y SYCL en CPU+GPU.

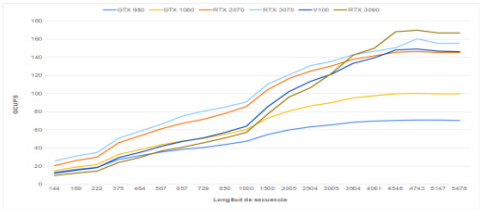
Conjunto de plataformas (H)	$\bar{\Phi}(\alpha, p, H)$ CUDA	SYCL
Intel CPU ∪ NVIDIA GPU	NA	27.39 %
Intel CPU ∪ AMD dGPU	NA	5.97 %
AMD CPU ∪ AMD iGPU	NA	14.03 %
Intel CPU ∪ Intel iGPU	NA	17.80 %
Intel CPU ∪ Intel dGPU	NA	14.49 %
Intel CPU ∪ Intel GPU	NA	14.57 %
Intel CPU∪ GPU	NA	22.51 %
CPU∪ GPU	NA	21.66 %



(a) Comparación de rendimientos al variar el tamaño del WG.



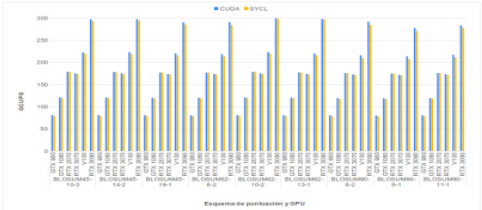
(b) Comparación de rendimientos al variar la base de datos de proteínas.



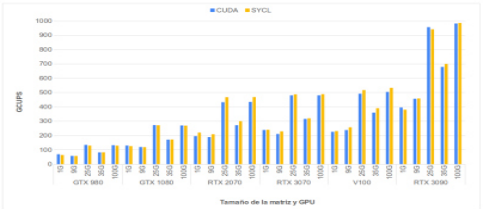
(c) Comparación de rendimientos al variar la longitud de las secuencias.



(d) Comparación de rendimiento al variar el algoritmo de alineamiento.



(e) Comparación de rendimiento al variar la matriz de puntuación.



(f) Comparación de rendimientos para alineamientos de ADN.

Figura 2: Comparación de rendimientos

Tabla 2: GCUPS y eficiencias arquitectónicas de los códigos CUDA y SYCL en GPUs individuales.

Plataforma			CUDA		SYCL	
Fabricante	GPU	GCUPS pico	GCUPS alcanzado	Eficiencia arq.	GCUPS alcanzado	Eficiencia arq.
NVIDIA	GTX 980	155.5	70.6	45.3 %	67.7	43.5 %
	GTX 1080	277.2	104.5	37.7 %	103.8	37.4 %
	RTX 2070	311.0	162.5	52.2 %	163.1	52.4 %
	Tesla V100	588.8	224.9	38.2 %	233.0	39.5 %
	RTX 3070	423.2	173.1	40.9 %	174.4	41.2 %
	RTX 3090	741.3	280.2	37.8 %	288.6	38.9 %
Intel	Arc A770	819.2	×	NA	191.4	23.3 %
	UHD 630	19.2	×	NA	13.1	68.4 %
	UHD 770	35.2	×	NA	26.6	75.7 %
AMD	RX 6700 XT	550.6	×	NA	284.4	51.7 %
	Vega 6	35.20	×	NA	7.5	21.3 %

Tabla 3: GCUPS y eficiencias arquitectónicas de los códigos CUDA y SYCL en múltiples GPUs.

Plataforma			CUDA		SYCL	
Fabricante	GPUs	GCUPS pico	GCUPS obt.	Arq efi.	GCUPS obt.	Arq efi.
NVIDIA	2× GTX 1080	554.6	189.8	34.2 %	187.8	33.9 %
	2× Tesla V100	846.4	318.1	27.0 %	336.5	28.6 %
	2× RTX 3070	1177.6	306.5	36.2 %	308.9	36.5 %
	Tesla V100 ∪ RTX 3090	1330.1	450.5	33.8 %	460.7	34.6 %
Intel	Arc A770 ∪ UHD 770	854.4	×	NA	126.8	14.8 %

Tabla 8: GCUPS y eficiencias arquitectónicas de los códigos CUDA y SYCL en CPUs individuales.

Plataforma			CUDA		SYCL	
Fabricante	CPUs	GCUPS pico	GCUPS obt.	Arq efi.	GCUPS obt.	Arq efi.
Intel	Core i5-7400	8.8	×	NA	4.2	47.6%
	Core i5-10400F	16.0	×	NA	8.2	51.2%
	Xeon E5-1620	9.3	×	NA	3.9	42.1%
	Xeon E5-2695	35.5	×	NA	15.8	44.6%
	Xeon Gold 6138	101.3	×	NA	28.5	28.2%
	Core i9-9900K	25.1	×	NA	9.6	38.4%
	Core i9-13900K	39.47	×	NA	15.9	40.2%
AMD	Ryzen 3	9.6	×	×	3.7	38.8%

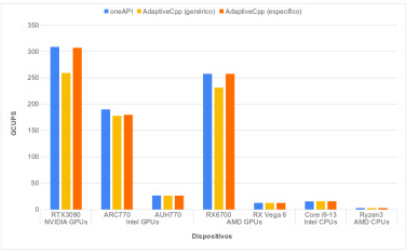


Tabla 9: Comparación de rendimiento entre implementaciones SYCL (oneAPI y AdaptiveCpp) en diferentes GPUs y CPUs de distintos proveedores.

Integración de métodos de descubrimiento de conocimiento embebido en fuentes de información desestructuradas

Autor: Juan M. Rodríguez (*jmrodriguez1982@gmail.com*)

Directores: Patricia Pesado, Rodolfo Bertone Asesor: Hernán Merlino

Universidad Nacional de La Plata, Facultad de Informática

Tesis defendida el lunes, 26 de junio de 2023

Introducción

En el año 2007 Michele Banko introduce un nuevo concepto en materia de extracción de información a partir de textos, al que llama en inglés: *Open Information Extraction* (Open IE o OIE). Se trata de un paradigma de extracción de conocimiento en donde un sistema informático realiza una sola pasada sobre el total de las fuentes de información no estructurada en formato de lenguaje natural (llamado corpus de documentos), dadas como entrada y extrae un gran conjunto de tuplas relacionales sin requerir ningún tipo de intervención humana. En el mismo trabajo Banko presenta un método llamado TEXTRUNNER, el cual es el primer método que trabaja dentro de este nuevo paradigma.

A partir de este trabajo se propusieron otros métodos de extracción de conocimiento bajo el paradigma que Banko llamó *Open Information Extraction* y que en español se lo puede llamar de forma más concreta: métodos de extracción de conocimiento para la Web, debido a que están pensados para trabajar con los grandes volúmenes de datos que provienen de Internet en general y de la Web en particular.

Para ilustrar lo anterior, considérese la oración del siguiente ejemplo:

Albert Einstein, quien nació en Ulm, ganó el Premio Nobel.

Extrayendo las relaciones semánticas presentes en la oración y expresándolas como una tupla en la forma: (Argumento 1, Relación, Argumento 2) obtenemos lo siguiente:

- (Albert Einstein, ganó el, Premio Nobel)
- (Albert Einstein, nació en, Ulm)

El propósito principal de los métodos de extracción de conocimiento para la Web, es hacer esta tarea de forma automática y eficiente.

Motivación

Construir un método de extracción de conocimiento para la Web que soporte idioma español y que sea capaz de extraer piezas de información con la misma efectividad con la que otros métodos en el estado del arte lo hacen para idioma inglés. Se proporcionará así una herramienta útil para el

procesamiento de lenguaje natural en español, mejorando la creación de bases de datos, sistemas de preguntas y respuestas, la elaboración de índices y catálogos, y la búsqueda y recuperación de información en documentos.

Problemas encontrados y solución propuesta

Los métodos existentes de extracción de conocimiento para la Web tienen una precisión y una exhaustividad (*recall*) considerablemente baja, de alrededor del 60%[1] y si bien tienen un gran potencial en cuanto a su aplicabilidad, es necesario mejorar su desempeño. Existen además otros problemas abiertos que están siendo abordados por varios autores como por ejemplo: la extracción de relaciones semánticas no informativas, la extracción de información subjetiva y el soporte para idiomas distintos del inglés.

El principal aporte de la tesis consistió en la publicación de tres métodos novedosos de extracción de conocimiento para la Web, uno para idioma inglés: ATP-OIE y dos para idioma español: TP-OIE-ES y ECMes. Así mismo, se propuso un marco de referencia único para la evaluación de los métodos, esto es la construcción de un conjunto de pruebas y una definición precisa de las métricas a utilizar y de cómo implementarlas.

ATP-OIE es un algoritmo autónomo, capaz de aprender de ejemplos y de aprender nuevos patrones de extracción mientras se está ejecutando de forma productiva. Por su parte, TP-OIE-ES replica el comportamiento de ATP-OIE para idioma español, con la salvedad de que no es capaz de aprender nuevos patrones mientras se ejecuta de forma productiva. Por último, ECMes es una versión reentrenada de TP-OIE-ES con otras mejoras adicionales. ECMes ha obtenido al momento de su publicación un mejor desempeño en idioma español, en los conjuntos evaluados, que otros métodos similares en el estado del arte.

Resultados obtenidos y aportes:

En esta sección se muestran tablas comparativas de los diversos métodos creados junto a otros modelos en el estado del arte.

Tabla 1. Medidas calculadas para ATP-OIE, ClausIE, OLLIE, ReVerb y MinIEen el conjunto de datos Reuters-103 propuesto en la tesis.

Métodos	Precisión	Exhaustividad	Medida F1
ClausIE	0.467	0.519	0.492
OLLIE	0.456	0.416	0.435
ReVerb	0.633	0.319	0.424
MinIE-C	0.612	**0.593**	**0.6022**
ATP-OIE autónomo	0,650	0,294	0,401

1 Estos valores corresponden al noviembre de 2022, antes de que aparecieran los grandes modelos de lenguaje (LLM) como GPT. Sí se consideraron los modelos de lenguaje intermedios como BERT, que tienen (y tenían) un rendimiento mucho menor al de los LLM. Incluso, para tareas especificas de extracción de información, estos modelos necesitaban ser combinados con métodos algorítmicos.

ATP-OIE asistido	**0,680**	0,401	0,504
ATP-OIE aprendizaje en-línea	0,670	0,390	0,493

Tabla 2. Medidas calculadas para TP-OIE-ES, DepOE y ArgOEutilizando como conjunto de datos 68 oraciones extraídas de textos escolares mexicanos.

Métodos	Precisión	Exhaustividad	Medida F1
TP-OIE-ES	0,62	**0,36**	**0,46**
DepOE	**0,89**	0.29	0,44
ArgOE	0,67	0,29	0,40

Tabla 3. Medidas calculadas para TP-OIE-ES, DepOE, ArgOE, ECMes y Multi^{2}OIE en el conjunto de 68 oraciones extraídas de textos escolares mexicanos.

Métodos	Precisión	Exhaustividad	Medida F1
TP-OIE-ES	0,62	0,36	0,46
ECMes	**0,92**	**0,42**	**0,57**
Multi^{2}OIE	0,46	0,19	0,27
DepOE	0,89	0.29	0,44
ArgOE	0,67	0,29	0,40

Tabla 4. Medidas calculadas para DepOE, ArgOE, ECMes y Multi^{2}OIE en un conjunto de textos en español propuesto en el trabajo de tesis.

Métodos	Precisión	Exhaustividad	Medida F1
ECMes	0,68	**0,34**	**0,45**
Multi^{2}OIE	0,22	0,10	0,14
DepOE	**0,81**	0,18	0,30
ArgOE	0,68	0,22	0,33

En la figura 1 se muestra de forma comparativa la medida F1 obtenida por ECMes y por otros modelos en el estado del arte:

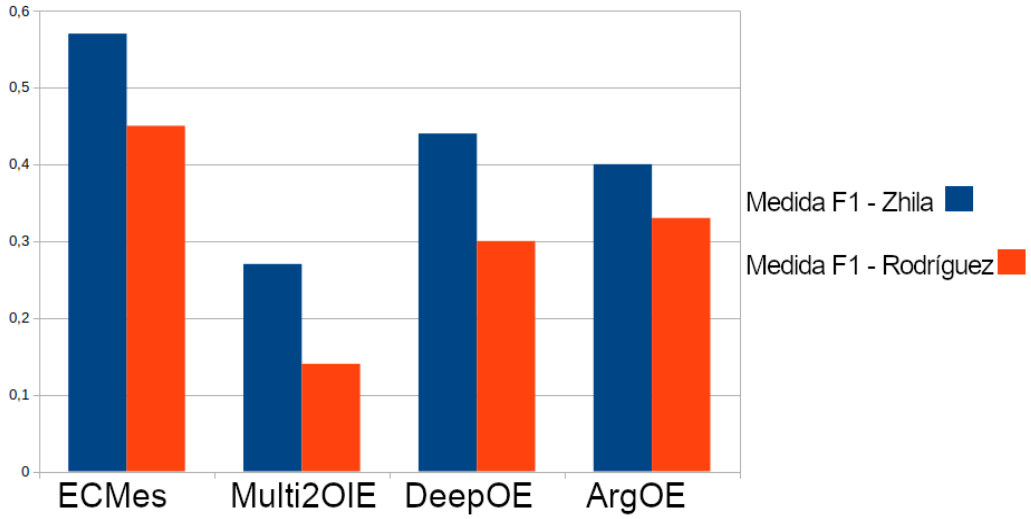


Figura 1: Medida F1 de los distintos métodos en ambos conjuntos de prueba

ECMes superó a TP-OIE-ES y a los otros métodos evaluados en idioma español, demostrando que la mecánica propuesta en TP-OIE funciona. Se puede concluir que ECMes es, hoy por hoy[2], un método de extracción de conocimiento para Web en el estado del arte para idioma español y es el que mejor desempeño tiene en los conjuntos de pruebas utilizados.

2 Al 1 de noviembre de 2022

Futuras líneas de investigación:

La tesis, entregada para su evaluación final el día 1 de noviembre de 2022 concluía que:

En paralelo a la proliferación de métodos de extracción de conocimiento para la Web, han aparecido algunos modelos puntuales para intentar crear una inteligencia artificial capaz de "comprender" texto en lenguaje natural o bien acercarse lo más posible a este fin. En particular, se trata de modelos de redes neuronales artificiales, preentrenados. Una de ellas es GPT-3 en 2020 y la otra es BERT en 2019. A pesar lo promisorio de estos modelos, no logran procesar el lenguaje natural al mismo nivel que un ser humano. Incluso uno de los métodos aquí analizados: Multi^{2}OIE, utiliza internamente BERT y aun así no es el método con mejor desempeño. De lograrse el objetivo de construir una IA capaz de comprender un texto en lenguaje natural de la misma forma en la que lo hace una persona, los métodos de extracción de conocimiento para la Web dejarían de tener sentido, ya que su funcionalidad sería sólo una de las muchas tareas factibles para estos sistemas. *Por lo cual, los problemas planteados seguirán abiertos hasta que pueda resolverse de forma satisfactoria la compresión de textos de forma automática*. Sin embargo, siempre hay lugar para mejorar, aunque sea mínimamente, los resultados actuales.

El 30 de noviembre de 2022 OpenAI lanzó chatGPT, una aplicación web para interactuar mediante mensajes de texto con un modelo nuevo llamado GPT-3.5 en el cual venía trabajando y que supera ampliamente a su predecesor GPT-3. A partir de entonces este y otros modelos de lenguaje han dejado obsoletos a los métodos de extracción de conocimiento. Cualquier futura línea de investigación para los algoritmos de extracción de conocimiento será una construcción que utilice como base un modelo de LLM. Tal que el algoritmo pueda refinar o validar las tuplas extraídas o ayudar en la creación de mapas conceptuales del conocimiento.

Método de Auditoría Informática basado en sistemas de procesamiento avanzado de datos para minimizar el riesgo de calidad de los resultados

Autor: Daisy Elizabeth Imbaquingo Esparza

Director: Javier Díaz

Institución: Facultad de Informática de la Universidad Nacional de La Plata, Argentina

Fecha de Exposición: 8 de septiembre de 2023

Resumen

En este documento se aborda la problemática de la auditoría informática en Instituciones de Educación Superior (IES), proponiendo un método innovador (MAIIES) diseñado para minimizar los riesgos asociados a la calidad y seguridad de los resultados de las auditorías. En las IES, existen estudios de propuestas de guías de auditoría informática aplicando metodologías, pero ninguna de ellas contempla los servicios y los procesos específicos que se desarrollan dentro de estas instituciones, con el objetivo de controlar y garantizar la seguridad de los activos tecnológicos frente a las diferentes amenazas e incidentes, así como determinar las oportunidades de mejora. Este método incluye la fase de planeación, ejecución, comunicación de resultados, validación y seguimiento del ejercicio de auditoría con cuarenta y siete actividades agrupadas en cada fase el mismo que busca optimizar la evaluación y acreditación institucional mediante la mejora de la eficacia y confiabilidad de las auditorías informáticas.

Introducción

La auditoría informática es un componente esencial en la gestión y el control de los recursos tecnológicos de una organización. En el ámbito de las IES, donde la infraestructura tecnológica soporta una amplia gama de servicios críticos, desde la gestión académica hasta la investigación, la auditoría informática desempeña un papel crucial en la garantía de la seguridad, la integridad y la eficiencia de los sistemas de información. Sin embargo, las metodologías de auditoría existentes a menudo no se adaptan a las particularidades de las IES, lo que puede comprometer la calidad de los resultados y limitar su utilidad para la toma de decisiones estratégicas. Este trabajo propone un método especializado que considera las necesidades y los desafíos específicos de las IES, para mejorar la precisión y la exhaustividad en los resultados de las auditorías.

Motivación

La principal motivación detrás de esta tesis es la identificación de una brecha significativa en las prácticas de auditoría informática en las IES. Los métodos tradicionales no incorporan técnicas estratégicas avanzadas ni se adaptan a los servicios y procesos específicos que se desarrollan en estas instituciones. Esta carencia puede conducir a riesgos sustanciales en la calidad y seguridad de los resultados de las auditorías, lo que a su vez limita su capacidad para contribuir a la evaluación y acreditación institucional. La presente investigación busca llenar este vacío mediante el desarrollo de

un método de auditoría informática que esté específicamente diseñado para abordar los desafíos únicos que enfrentan las IES en la gestión de sus recursos tecnológicos.

Aporte

El principal aporte de esta tesis es el desarrollo y la estandarización del MAIIES (Método de Auditoría Informática para Instituciones de Educación Superior). Este método proporciona una guía estructurada que abarca las fases de planificación, ejecución, comunicación de resultados, validación y seguimiento. El MAIIES se compone de cuarenta y siete actividades, cada una de ellas diseñadas para controlar y garantizar la seguridad de los activos tecnológicos, identificar amenazas e incidentes, y determinar oportunidades de mejora en las IES. El método propuesto se basa en un enfoque integral que combina técnicas de auditoría tradicionales con herramientas de procesamiento avanzado de datos, lo que permite una evaluación más exhaustiva y precisa de los riesgos y las vulnerabilidades en los sistemas de información de las IES.

Metodología

La investigación se llevó a cabo en dos fases principales. En la primera fase, se realizó una extensa revisión bibliográfica, la identificación de factores y métricas de calidad y seguridad de la información en auditorías, así como para analizar los marcos de referencia existentes. También se llevó a cabo un estudio de la situación actual en las IES de Ecuador que pertenecen a la Corporación Ecuatoriana para el Desarrollo de la Investigación y la Academia (CEDIA). En la segunda fase, se recomendó el diseño y la estandarización del MAIIES, basado en los hallazgos de la primera fase y en las mejores prácticas en el campo de la auditoría informática.

Factores que Impactan la Calidad de la Auditoría

La calidad de la auditoría informática se ve influenciada por diversos factores que pueden clasificarse en tres categorías principales:

1. **Factor Humano:** Este factor se refiere a las competencias, la experiencia y la ética profesional de los auditores. La capacitación adecuada, la actualización constante de conocimientos y la independencia del auditor son cruciales para garantizar la calidad de la auditoría.
2. **Factor Técnico:** Este factor se relaciona con las herramientas, las técnicas y las metodologías utilizadas en la auditoría. La disponibilidad de software especializado, la aplicación de estándares internacionales y la adopción de enfoques basados en riesgos son elementos clave para mejorar la calidad de la auditoría.
3. **Factor Contextual o del Entorno:** Este factor considera el entorno organizacional, las políticas de seguridad de la información y la cultura de cumplimiento de las IES. Un entorno favorable, con un fuerte compromiso de la alta dirección y una cultura de seguridad arraigada, facilita la realización de auditorías de alta calidad.

Métricas para Evaluar la Calidad y Seguridad de la Información

La evaluación de la calidad y seguridad de la información en las auditorías informáticas se basa en el uso de métricas específicas que permiten medir el nivel de cumplimiento de los objetivos y la efectividad de los controles implementados.

Algunas de las métricas clave incluyen:

- **Tasa de detección de vulnerabilidades:** Mide la capacidad de la auditoría para identificar vulnerabilidades en los sistemas de información.
- **Tiempo de respuesta ante incidentes:** Evalúa la eficiencia de la IES para responder a incidentes de seguridad.
- **Nivel de cumplimiento de las políticas de seguridad:** Mide el grado en que la IES cumple con sus propias políticas de seguridad de la información.
- **Satisfacción de los usuarios:** Evalúa la percepción de los usuarios sobre la calidad y la seguridad de los servicios de información.

Con las 91 métricas, se llevó a cabo un análisis centrado específicamente en auditorías informáticas dentro de las IES (Instituciones de Educación Superior), obteniendo respuestas de 54 IES en Ecuador que habían implementado previamente algún tipo de auditoría informática, logrando así una visión de los requisitos para que una auditoría sea de calidad y asegure la información en todo momento. Finalmente, se obtuvieron 42 métricas de calidad y 18 de seguridad, que forman parte de los instrumentos de evaluación y validación del MAIIES.

Resultados

El resultado principal de esta investigación es un método de auditoría informática que se adapta a las necesidades específicas de las IES. El método incluye una serie de herramientas y técnicas que permiten a los auditores evaluar de manera integral la seguridad y la calidad de los sistemas de información. Además, el MAIIES proporciona un marco de referencia para la mejora continua de las prácticas de auditoría, lo que permite a las IES mantenerse al día con los últimos avances en tecnología y seguridad de la información.

A continuación, se presenta una tabla resumen de los resultados de la aplicación del MAIIES en una IES piloto:

Tabla 1: Resultados de la Aplicación del MAIIES en la IES Piloto

Métrica	Valor Inicial	Valor Final	Mejora (%)
Tasa de detección de vulnerabilidades	60%	90%	50%
Tiempo de respuesta ante incidentes (horas)	24	8	66%
Nivel de cumplimiento de políticas	70%	95%	35%

Métrica	Valor Inicial	Valor Final	Mejora (%)
Satisfacción de los usuarios (escala 1-5)	3	4.5	50%

Además, se incluye el siguiente gráfico que muestra la mejora en la tasa de detección de vulnerabilidades tras la aplicación del MAIIES:

Al aplicar el método propuesto, las IES pueden mejorar sus procesos, ya que un ejercicio de auditoría exitoso identifica debilidades y fallos para proporcionar recomendaciones que apoyan la toma de decisiones y la mejora continua del proceso dentro del área de TI, beneficiando a toda la organización. Además, a medida que una universidad alcanza objetivos genéricos y específicos en un nivel de madurez, incrementa su madurez y, simultáneamente, logra el cumplimiento de las leyes y regulaciones nacionales relevantes.

Disciplina

La tesis se enmarca dentro del campo de la auditoría informática, con un enfoque específico en la aplicación de técnicas de procesamiento avanzado de datos para mejorar la calidad y la seguridad de los procesos de auditoría en las Instituciones de Educación Superior. La investigación aborda temas relevantes en la seguridad de la información, la gestión de riesgos y la evaluación de la conformidad, y contribuye al desarrollo de mejores prácticas en la auditoría informática en el contexto de las IES.

Discusión

El MAIIES representa un avance significativo en la auditoría informática en las IES, ya que proporciona un método estructurado y adaptable que se basa en las mejores prácticas y en las últimas tecnologías. Sin embargo, es importante reconocer que la implementación del MAIIES puede presentar desafíos, como la necesidad de capacitación del personal y la adaptación de los procesos existentes. Además, es fundamental que las IES adopten un enfoque integral de la seguridad de la información, que incluya políticas, procedimientos y controles que complementen el MAIIES.

Conclusiones

La tesis concluye que el MAIIES es un método eficaz para mejorar la calidad y la seguridad de las auditorías informáticas en las IES. El método proporciona una guía estructurada que abarca todas las fases del proceso de auditoría, desde la planificación hasta el seguimiento, y se basa en un enfoque

integral que combina técnicas de auditoría tradicionales con herramientas de procesamiento avanzado de datos. La implementación del MAIIES puede ayudar a las IES a garantizar la seguridad, la integridad y la eficiencia de sus sistemas de información, y a mejorar su capacidad para tomar decisiones estratégicas basadas en información precisa y confiable.

Posibles Líneas de Investigación Futura

- **Implementación y Evaluación Empírica:** Aplicar el MAIIES en diversas IES y evaluar su efectividad en la mejora de la calidad y seguridad de las auditorías informáticas.
- **Integración de Inteligencia Artificial:** Investigar la incorporación de técnicas de inteligencia artificial y aprendizaje automático para automatizar y mejorar la eficiencia de las auditorías.
- **Adaptación a Otros Sectores:** Adaptar el MAIIES para su aplicación en otros sectores que requieran auditorías informáticas especializadas, como el sector de la salud o el sector financiero.
- **Desarrollo de Herramientas de Software:** Crear herramientas de software que faciliten la implementación y el seguimiento del MAIIES en las IES.
- **Estudio de Factores Humanos:** Profundizar en el estudio de los factores humanos que influyen en la calidad de las auditorías informáticas, como la capacitación y la experiencia de los auditores.
- **Análisis Comparativo de Marcos de Referencia:** Realizar un análisis comparativo exhaustivo de los diferentes marcos de referencia utilizados en la auditoría informática, con el fin de identificar las mejores prácticas y las áreas de mejora.
- **Desarrollo de Métricas de Desempeño:** Desarrollar métricas de desempeño que permitan evaluar la efectividad del MAIIES en la mejora de la seguridad y la calidad de los sistemas de información en las IES.
- **Investigación sobre la Percepción de los Auditores:** Investigar la percepción de los auditores sobre el MAIIES y su utilidad en la mejora de su trabajo.
- **Estudio de Caso en IES:** Realizar un estudio de caso en una o varias IES para evaluar la implementación del MAIIES y sus resultados en la práctica.
- **Desarrollo de un Modelo de Madurez:** Desarrollar un modelo de madurez que permita a las IES evaluar su nivel de preparación para la implementación del MAIIES y establecer un plan de mejora continua.

Bibliografía

- Díaz, J. (2018). *Auditoría Informática: Un Enfoque Práctico*. Buenos Aires: Alfaomega.
- ISACA. (2018). *COBIT 2019 Framework: Governance and Management Objectives*. Rolling Meadows, IL: ISACA.

- ISO/IEC 27001. (2013). *Information technology — Security techniques — Information security management systems — Requirements*. Geneva: International Organization for Standardization.
- IT Governance Institute. (2007). *ITIL V3*.
- Ron, M. (2015). *Seguridad de la Información en las Organizaciones*. Quito: Editorial Universitaria.
- Schneier, B. (2004). *Secrets and Lies: Digital Security in a Networked World*. Indianapolis: Wiley.

Metodología basada en Modelos en Tiempo de Ejecución para la Construcción y Operación de Sistemas Autoconscientes de Internet de las Cosas

Universidad Nacional de La Plata – Facultad de Informática

Autor: Dr. Lenin Erazo Garzón
Directora: Dra. Priscila Cedillo Orellana
Codirector: Dr. Gustavo Rossi

Fecha de defensa: 16 de diciembre de 2024

1. Problemática / motivación

El Internet de las Cosas (IoT) es un paradigma tecnológico utilizado para referirse a una infraestructura compuesta por un conjunto de objetos (cosas) inteligentes, ubicuos e interconectados mediante las tecnologías de información y comunicación para procesar datos del mundo físico y virtual; y, actuar con la mínima intervención humana [1][2]. El IoT promueve el desarrollo de innovadoras aplicaciones en una diversidad de dominios (p. ej., hogar, salud, ciudad inteligente, industria), agregando una nueva dimensión a la vida de las personas, centrada en la interacción permanente con los objetos inteligentes, en cualquier lugar y momento, lo cual impacta favorablemente en el bienestar y desarrollo sostenible de la sociedad [3]. Según Gartner [4], el número total de dispositivos IoT instalados en el año 2017 (8.400 millones), superó a la población total de seres humanos; mientras que, una previsión de la International Data Corporation (IDC) [5] estima que habrán alrededor de 41.600 millones de dispositivos IoT instalados en el 2025.

En la actualidad, uno de los desafíos más importantes para la comunidad científica y la industria es el desarrollo de software para sistemas IoT, dado que se caracterizan por ser altamente distribuidos, heterogéneos, escalables, dinámicos e inciertos [6][7]. En particular, los sistemas IoT operan en entornos ubicuos que cambian constantemente, debiendo ser controlados oportunamente para garantizar su correcto y continuo funcionamiento [6]. Debido a la incertidumbre intrínseca asociada a los cambios, resulta complejo identificar todos los requisitos de un sistema IoT durante la etapa de desarrollo [8]. Además, las metodologías y herramientas de ingeniería de software tradicionales han demostrado ser poco efectivas para afrontar el dinamismo y la escalabilidad de los sistemas IoT, siendo los desarrolladores los que gestionan la evolución de estos sistemas en el tiempo de diseño [9][10]. Esta limitación se agudiza aún más, ya que estos enfoques metodológicos y herramientas se centran en los detalles de implementación (programación a bajo nivel) en lugar del dominio del problema, volviéndose una tarea compleja y tediosa para los desarrolladores por la heterogeneidad de los nodos de computación y dispositivos IoT (sensores, actuadores) que deben ser integrados, a nivel de las capas Edge, Fog y Cloud de una infraestructura IoT [11].

Un enfoque apropiado para abordar los problemas señalados es incorporar la autoconsciencia computacional a los sistemas IoT con el apoyo de la Ingeniería Dirigida por Modelos (MDE) y, particularmente los modelos en tiempo de ejecución (models@run.time), a fin de aumentar su capacidad autónoma y resiliente frente a las situaciones imprevistas [12][13]. Un sistema IoT es considerado autoconsciente, si es capaz de generar conocimiento sobre sí mismo y el entorno (p. ej., estructura, estado, evolución), sin un control externo [14]. Este conocimiento puede ser capturado directamente de los datos crudos recolectados desde los sensores físicos o virtuales (procesos pre-reflexivos) o extraído de modelos conceptuales internos, construidos a partir de datos históricos (procesos reflexivos) [15].

El término autoconsciencia emerge desde la psicología y los primeros intentos por abordar

explícitamente la autoconsciencia, en un contexto informático, comenzaron alrededor del año 2004 [8]. A su vez, la computación autoconsciente abarca varios aspectos que han sido estudiados implícitamente por otras áreas de investigación de la informática (p. ej., inteligencia artificial, sistemas multiagentes, computación autónoma, sistemas autoadaptativos, computación reflexiva, modelos en tiempo de ejecución); sin embargo, el área de los sistemas informáticos autoconscientes todavía es poco explorada; y, plantea importantes desafíos y oportunidades para la comunidad científica [16]. Estos desafíos están enfocados en trasladar varias de las funcionalidades de los sistemas que tradicionalmente se programan en tiempo de diseño al tiempo de ejecución [17], a través de la búsqueda de soluciones que incorporen el aprendizaje y razonamiento (capacidades de autoconsciencia) en línea basado en modelos; y, que a su vez sean viables técnica y económicamente, con respecto a la cantidad de recursos de hardware y software que necesitan para apoyar los grados de autoconsciencia de un sistema [16].

Particularmente, Elhabbash et al. [18] en su revisión sistemática sobre la adopción de la autoconsciencia en la ingeniería de software, evidencian que los trabajos de investigación sobre autoconsciencia dentro del dominio de IoT son escasos. Los esfuerzos de investigación se han concentrado fundamentalmente en trabajos teóricos orientados a establecer hojas ruta para la adopción de la autoconsciencia en sistemas IoT, así como soluciones prácticas a medida (ad hoc) para aplicaciones particulares de IoT, descuidando la concepción de metodologías e infraestructuras genéricas e integrales para la construcción de sistemas IoT con características autoconscientes.

Por su parte, la Ingeniería Dirigida por Modelos (Model-Driven Engineering - MDE) es un paradigma que ha incrementado la importancia de los modelos, ya que no solo se utilizan con fines de documentación y comunicación, sino como artefactos centrales del proceso de desarrollo de software [19][20]. MDE propone enfoques de desarrollo de software en los que se crean modelos abstractos de sistemas de software y se transforman sistemáticamente en implementaciones concretas (p. ej., código fuente, archivos de configuración) [21]. Además, en los últimos años, varias iniciativas de investigación han llevado la idea de utilizar los modelos, un paso más adelante, en tiempo de ejecución (models@run.time) [22], demostrando su utilidad para hacer frente a los aspectos complejos, dinámicos e inciertos de los sistemas modernos (p. ej., IoT) y su entorno [23][24][25]. Los modelos en tiempo de ejecución son una autorrepresentación causalmente conectada de un sistema asociado que enfatiza la estructura, el comportamiento y/o los objetivos del sistema y que puede ser manipulada en tiempo de ejecución para fines específicos [23][26].

En el ámbito de la autoconsciencia computacional, los modelos en tiempo de ejecución son el núcleo de los sistemas autoconscientes y proporcionan el medio para la autorrepresentación del sujeto (sistema) que es consciente [16]. Es así que, la información recopilada desde el sistema observado puede ser utilizada para retroalimentar los modelos en tiempo de ejecución. Luego, por medio de estos modelos se puede analizar el estado y la posible evolución del sistema; y, en caso de ser necesario recomendar acciones correctivas [27].

Según la revisión sistemática de Bencomo et al. [25] sobre los modelos en tiempo de ejecución, uno de los desafíos más relevantes de los modelos en tiempo de ejecución es la autoconsciencia computacional, debido a los pocos trabajos encontrados en la literatura, existiendo una importante brecha de investigación a ser aprovechada. Puntualmente, los autores consideran prioritario enfocarse en la aplicación de los modelos en tiempo de ejecución para construir sistemas autoconscientes de su entorno y de la interacción con otros sistemas o entre sus subsistemas o componentes (autoconsciencia colectiva), siendo las principales preocupaciones de los sistemas IoT.

Frente a esta problemática, la tesis doctoral propone una metodología (denominada Aware-IoT@Run.Time) e infraestructura de soporte basada en modelos en tiempo de ejecución que apoye a los stakeholders (p. ej., desarrolladores, expertos de dominio) en la construcción, operación y

mantenimiento de sistemas IoT autoconscientes, con un alto grado de abstracción, automatización y flexibilidad. A su vez, la solución tecnológica propuesta está fundamentada en estándares, frameworks conceptuales, arquitecturas de referencia y ontologías relevantes sobre IoT, autoconsciencia computacional y medición de la calidad de software, lo cual le otorga un sustento teórico, formalidad, rigor científico y coherencia semántica a la misma.

La estructura de este artículo es la siguiente: Las secciones 2 y 3 presentan una visión de general del proceso metodológico de Aware-IoT@Run.Time y los componentes de su infraestructura de soporte, respectivamente. La sección 4 describe brevemente la evaluación empírica de la solución tecnológica. La sección 5 detalla las principales contribuciones de la tesis doctoral. Finalmente, en la sección 6 se discuten las líneas de investigación futuras.

2. Metodología Aware-IoT@Run.Time

La metodología propuesta Aware-IoT@Run.Time [27][28] está conformado por un conjunto de actividades, recursos, artefactos y herramientas que indican cómo construir, operar y mantener un sistema IoT autoconsciente. Durante el trabajo de investigación, esta metodología ha sido objeto de varias evaluaciones empíricas que han conllevado a su refinamiento y mejora continua, a fin de garantizar un proceso sistemático, ordenado, eficaz, eficiente y repetible.

La metodología está construida a partir del bucle de aprendizaje y razonamiento basado en modelos, por sus siglas en inglés LRA-M loop [14], el cual describe los principales procesos de un sistema informático autoconsciente, iniciando desde la observación directa para recopilar datos empíricos sobre el propio sistema y su entorno. Luego, se producen procesos de aprendizaje continuo para capturar el conocimiento sobre los aspectos relevantes del sistema IoT y su entorno, por medio de la creación y/o aplicación de modelos conceptuales o reflexivos que utilizan el historial de los datos empíricos recolectados u otros datos derivados a partir de estos. Finalmente, se ejecutan procesos de razonamiento sobre el conocimiento adquirido para comprender cuantitativa y cualitativamente el estado actual y la evolución del sistema. Según LRA-M, un sistema autoconsciente no está obligado a actuar (autoadaptar), por lo que la metodología propuesta llega hasta determinar las recomendaciones o acciones que deben realizarse para asegurar o apoyar la consecución de los objetivos del sistema IoT.

A su vez, varios recursos y artefactos que forman parte de la metodología han sido construidos a partir de estándares o trabajos de investigación relevantes, entre los más importantes: la arquitectura de referencia para IoT ISO/IEC 30141:2018 [1], el framework conceptual de autoconsciencia de Lewis et al. [15] y la ontología de medición de software de García et al. [29].

En la Figura 1 se presenta un diagrama de proceso de alto nivel basado en la notación SPEM 2.0 (Software & Systems Process Engineering Meta-Model Specification) [30] que muestra el orden de ejecución de las actividades que conforman la metodología, así como los roles de usuario, herramientas y artefactos de entrada y salida que se utilizan durante el proceso. La metodología está compuesta por seis actividades principales. Las dos primeras actividades requieren la intervención de los desarrolladores; mientras que, las cuatro restantes se ejecutan de manera automática con el apoyo de las herramientas de software propias de la infraestructura de soporte.

Desde un ámbito general, el proceso metodológico propuesto comienza con dos actividades de modelado. La *primera actividad* está orientada al diseño visual del modelo de arquitectura de un sistema IoT que soporte los procesos de monitoreo, a nivel de las capas Edge, Fog y Cloud [31]. Mientras que, la *segunda actividad* se enfoca en la construcción de un modelo con especificaciones de alto nivel sobre las capacidades de autoconsciencia distribuidas de un sistema IoT.

Por su parte, la *tercera actividad* utiliza los modelos de arquitectura y autoconsciencia en tiempo de ejecución para implementar de forma automática y distribuida los recursos de software que soportan

la infraestructura de monitoreo y autoconsciencia del sistema IoT. Además, esta actividad ante cualquier cambio en los modelos en tiempo de ejecución es capaz de realizar reconfiguraciones sobre los recursos de software; es decir, mantiene la relación causal entre los modelos y el sistema IoT observado.

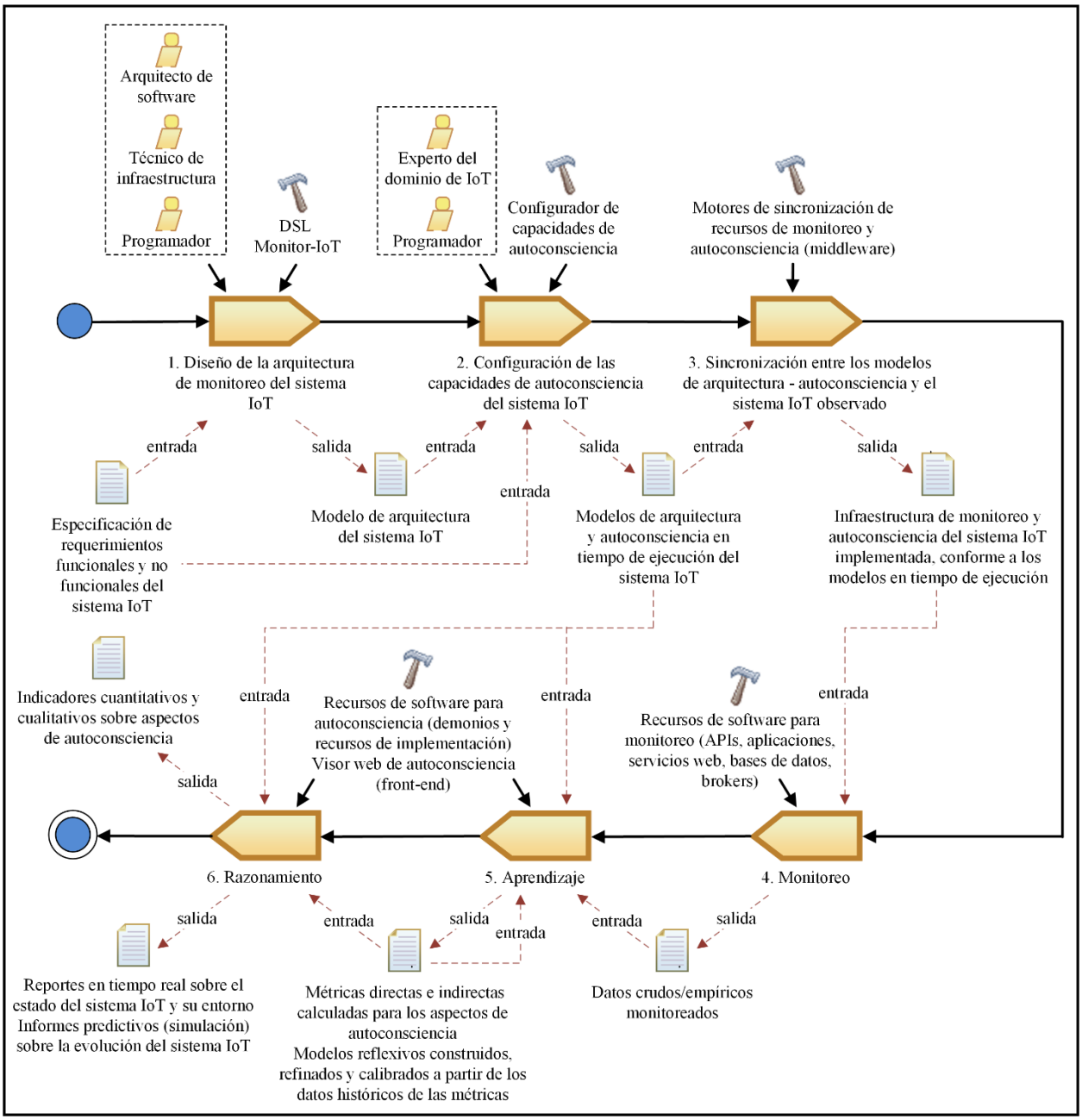


Figura 1. Proceso metodológico de alto nivel de Aware-IoT@Run.Time [27][28].

Las restantes actividades de la metodología utilizan los recursos de software de monitoreo y autoconsciencia implementados y puestos en operación. Estos recursos se encargan de recolectar, calcular y calibrar en tiempo de ejecución los parámetros (métricas, metadatos, etc.) especificados en los modelos, a fin de que reflejen en todo momento el estado y la evolución del sistema IoT en operación, cerrando de esta forma, el ciclo de la relación causal entre los modelos y el sistema IoT asociado. En particular, *la cuarta actividad* emplea los recursos de monitoreo para recolectar transportar y procesar los datos crudos o empíricos de los sensores u otras fuentes; y, consecuentemente almacenarlos en las bases de datos previamente especificadas en el modelo de arquitectura. Por su parte, la *quinta actividad* utiliza los datos empíricos recopilados previamente y

los recursos de autoconsciencia para producir las métricas directas e indirectas que valoran los aspectos de autoconsciencia de un sistema IoT. Finalmente, la *sexta actividad*, asimismo con el apoyo de los recursos de autoconsciencia, transforma las métricas directas e indirectas en indicadores cuantitativos y cualitativos, a fin de que el sistema IoT a partir de estos, sea capaz de conocer su estado y posible evolución futura, así como proponer recomendaciones o acciones correctivas de ser necesario.

3. Infraestructura de soporte a la metodología Aware-IoT@Run.Time

Todas las actividades de la metodología Aware-IoT@Run.Time están soportadas por una infraestructura de software conformada por un conjunto de componentes (artefactos o herramientas). La Figura 2 presenta la arquitectura de la infraestructura de soporte, incluyendo los componentes y sus interrelaciones. Estos componentes se enumeran a continuación [27]:

1) Metamodelo de arquitectura de monitoreo de sistemas IoT (Monitor-IoT) [32].
2) Metamodelo de capacidades de autoconsciencia de sistemas IoT (Aware-IoT).
3) Lenguaje específico de dominio (DSL) Monitor-IoT [32].
4) Configurador web de capacidades de autoconsciencia para sistemas IoT.
5) Middleware de la infraestructura, conformado por los motores de sincronización de recursos de software de monitoreo y autoconsciencia para sistema IoT.
6) Visor web de métricas de autoconsciencia.

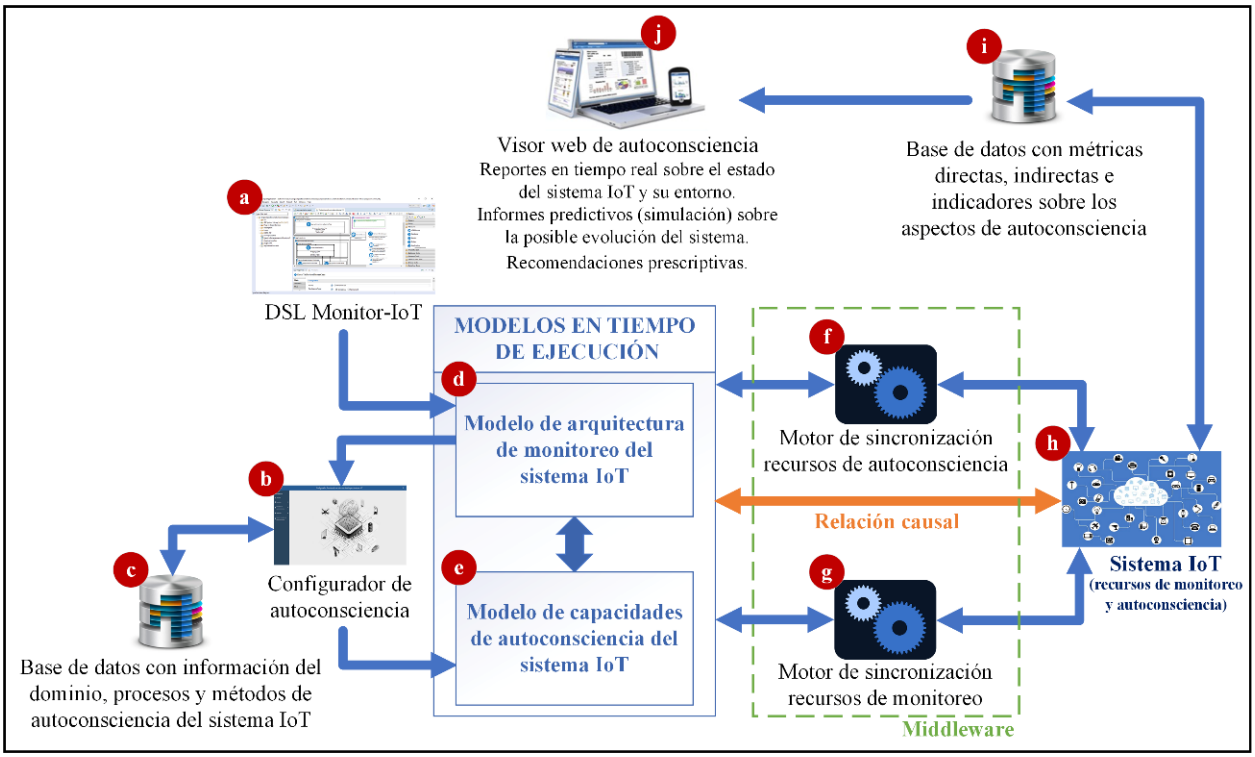


Figura 2. Arquitectura general de la infraestructura de soporte basada en models@run.time [27].

En un contexto general, las herramientas *DSL Monitor-IoT* y *configurador web de autoconsciencia* apoyan a los usuarios en la realización de las dos primeras actividades de la metodología, las cuales tienen como propósito la construcción de los modelos de arquitectura y autoconsciencia de un sistema IoT, respectivamente. Estas herramientas utilizan la sintaxis y semántica contenida en los *metamodelos Monitor-IoT* y *Aware-IoT*; de ahí que, los modelos de arquitectura y autoconsciencia resultantes son conformes a estos metamodelos. A su vez, el *configurador web de autoconsciencia* utiliza como insumo el modelo de arquitectura creado en el *DSL Monitor-IoT* y dispone de una base de datos para almacenar tanto las configuraciones particulares de un sistema IoT, relacionadas con su

dominio, procesos y métodos de autoconsciencia, como las configuraciones generales sobre las métricas, criterios de decisión y recursos de implementación (fórmulas, funciones, servicios web) que pueden ser reutilizados en la construcción de diversos modelos de autoconsciencia para sistemas IoT (ver Fig. 2(a-e)).

Por su parte, los *motores del middleware de la infraestructura* tienen como alcance la automatización de la tercera actividad de la metodología. Por lo tanto, esto motores utilizan los modelos de arquitectura y autoconsciencia en tiempo de ejecución para implementar y/o reconfigurar automáticamente los artefactos y recursos de software que soportan los procesos de monitorización y autoconsciencia del sistema IoT observado, a nivel de las capas Edge, Fog y Cloud. De esta forma, cualquier cambio en los modelos se refleja automáticamente en el sistema IoT en operación, garantizando la relación de causalidad entre estos (ver Fig. 2(f-h)).

En particular, los artefactos de software implementados por el *motor de sincronización de recursos de monitoreo* incluyen APIs, aplicaciones, servicios web, tópicos de brokers y bases de datos, los cuales operan de manera interconectada para soportar la recolección, transporte, procesamiento y almacenamiento de los datos crudos de los sensores u otras fuentes. Mientras que, los artefactos de software implementados por el *motor de sincronización de recursos de autoconsciencia* incluyen demonios (aplicaciones que se ejecutan cada cierto intervalo de tiempo) que encapsulan y ejecutan de manera distribuida y en segundo plano los procesos pre-reflexivos y reflexivos, a fin calcular las métricas para evaluar los aspectos de autoconsciencia. Además, como parte de los artefactos de autoconsciencia se tienen las funciones y servicios web que son llamados o consumidos desde los demonios para operacionalizar los métodos de cálculo y modelos de análisis, así como las bases de datos que almacenan las métricas directas, indirectas e indicadores obtenidos por los procesos de autoconsciencia (ver Fig. 2(i)).

Por último, la infraestructura de soporte dispone de un *visor web de autoconsciencia*, a nivel de front-end, que presenta: i) reportes (tabular y gráfico) en tiempo real sobre el estado de los aspectos de autoconsciencia observados, incluyendo la valoración cuantitativa, cualitativa e interpretación de las mismas, así como las recomendaciones o acciones prescriptivas, en caso de existirlas; e, i) informes predictivos (simulación) sobre la posible evolución de un aspecto de autoconsciencia, considerando los escenarios, variables y valores de simulación disponibles en el modelo de autoconsciencia (ver Fig. 2(j)).

4. Evaluación empírica de la solución tecnológica

La metodología Aware-IoT@Run.Time y sus herramientas de soporte fueron evaluadas empíricamente mediante un estudio de caso y dos cuasi experimentos en el subdominio de ambientes de vida asistida. El estudio de caso demostró que siguiendo paso a paso la guía metodológica y con el apoyo de las herramientas es factible construir sistemas IoT autoconscientes con un alto grado de abstracción, automatización y flexibilidad. En este sentido, se obtuvieron resultados muy favorables, entre los cuales se destacan: i) el proceso metodológico de Aware-IoT@Run.Time orientó de manera formal, disciplinada y ágil al equipo de desarrollo; ii) las herramientas de la infraestructura permitieron que el equipo de desarrollo se concentre fundamentalmente en el dominio del problema (especificaciones de alto nivel), minimizando los detalles de implementación (programación a bajo nivel); iii) los metamodelos Monitor-IoT y Aware-IoT proporcionaron un vocabulario amplio para especificar la diversidad de requerimientos de monitoreo y autoconsciencia establecidos en el estudio de caso; y, iv) el middleware mantuvo correctamente la relación causal entre los modelos y el sistema IoT observado, y viceversa.

Por su parte, los cuasi experimentos basados en el Modelo de Evaluación de Métodos (MEM) [33] y ejecutados en un entorno académico determinaron la alta efectividad y eficiencia de los participantes al utilizar la solución Aware-IoT@Run.Time. Además, estos cuasi experimentos demostraron la

intención de los participantes de utilizar la solución a futuro, ya que la consideran fácil de usar y útil.

5. Principales aportes de la tesis doctoral

La presente tesis doctoral ha contribuido con una solución tecnológica que combina la autoconsciencia computacional y la Ingeniería Dirigida por Modelos (models@run.time) para hacer frente a los problemas y desafíos que atañen al desarrollo de software para sistemas IoT, producto de los escenarios altamente distribuidos, heterogéneos, escalables, e inciertos en los que operan. Entre las principales contribuciones se destacan las siguientes:

1. ***Estado del arte completo y actualizado sobre la autoconsciencia computacional en IoT y otros dominios relacionados (sistemas ciber físicos, ciudad inteligente, computación ubicua),*** con énfasis en aquellos estudios que emplean los paradigmas de MDE (models@run.time).
2. ***Metodología Aware-IoT@Run.Time.*** Enfoque genérico, aplicable a cualquier subdominio de IoT, que utiliza los modelos como artefactos centrales del proceso de desarrollo, operación y mantenimiento de sistemas IoT autoconscientes. Particularmente, los sistemas IoT construidos bajo este enfoque, utilizan los modelos en tiempo de ejecución para comprender el estado y la posible evolución de sí mismo y su entorno; y, determinar las respectivas recomendaciones de ser necesario. Entre las principales fortalezas, Aware-IoT@Run.Time abstrae la complejidad de los sistemas IoT, eleva el nivel de automatización del desarrollo, mejora la sincronización entre los artefactos de diseño e implementación de los sistemas IoT, minimiza los detalles de implementación (programación a bajo nivel) y mejora la gestión de la evolución de estos sistemas (incrementa su autonomía y resiliencia).
3. ***Infraestructura de soporte a la metodología Aware-IoT@Run.Time.*** Conjunto de artefactos y herramientas de software que cubren todo el proceso metodológico de Aware-IoT@Run.Time y viabilizan la construcción, operación y mantenimiento de sistemas IoT autoconscientes, con un alto grado de abstracción, automatización y flexibilidad.

3.1. DSL Monitor-IoT. Lenguaje específico de dominio visual de alto nivel para diseñar arquitecturas de sistemas IoT que soporten el monitoreo, a través de las capas de Edge, Fog y Cloud. Monitor-IoT está conformado por un diseñador gráfico (sintaxis concreta) construido en las herramientas Obeo Designer Community y Eclipse Sirius, que simplifica y agiliza las tareas de creación y mantenimiento de modelos de arquitecturas de IoT. Además, Monitor-IoT está basado en un metamodelo (sintaxis abstracta) especificado en Ecore y alineado con la arquitectura de referencia para IoT ISO/IEC 30141:2018 [1], proporcionando un lenguaje más completo capaz de modelar una amplia variedad de entidades digitales y flujos de datos (síncronos y asíncronos) entre estas, a nivel de las capas Edge, Fog y Cloud; y, de esta forma soportar el monitoreo en una diversidad de escenarios de IoT. Principalmente, Monitor-IoT provee un vocabulario para definir: i) las entidades físicas y digitales a ser monitoreadas; ii) las entidades digitales (nodos de computación y sus recursos) que soportan los procesos de monitoreo (recolección, transporte, procesamiento y almacenamiento de datos), a nivel de las capas Edge, Fog y Cloud; iii) las propiedades a monitorear para cada entidad IoT; iv) los flujos de datos entre las entidades digitales, basados en comunicaciones síncronas o asíncronas; y, v) las operaciones de agregación a ser aplicadas sobre los datos recolectados.

3.2. Configurador de autoconsciencia para sistemas IoT. Aplicación web desarrollada en Node.js para construir modelos con especificaciones de alto nivel sobre una amplia variedad de niveles y aspectos de autoconsciencia para sistemas IoT. Estas especificaciones son expresadas mediante conceptos abstractos del dominio de la autoconsciencia computacional; y, complementados con conceptos sobre medición de la calidad de software. Para ello, el configurador está basado en el metamodelo Aware-IoT, construido a partir del framework conceptual de autoconsciencia de Lewis et al. [15] y la ontología de medición de software de García et al. [29]; e, integrado con el

metamodelo Monitor-IoT. Específicamente, el configurador web permite definir: i) el dominio (sujeto, objeto) de autoconsciencia de un sistema IoT; ii) los aspectos de autoconsciencia a capturar de las entidades IoT, pudiendo ser individuales o colectivos; iii) los procesos de autoconsciencia pre-reflexivos y reflexivos, junto con sus métodos de autoconsciencia (recolección, cálculo y modelos de análisis) y recursos de implementación (fórmulas, funciones, servicios web); iv) las métricas (directas, indirectas e indicadores) para evaluar los aspectos de autoconsciencia; y, v) los criterios de decisión, umbrales y recomendaciones prescriptivas asociados a los modelos de análisis.

3.3. Middleware. Aplicación ligera y multiplataforma para mantener la relación causal entre los modelos de arquitectura - autoconsciencia y el sistema IoT asociado, y viceversa. Este middleware ante cualquier cambio en las especificaciones de los modelos, implementa o reconfigura de forma automática y distribuida los artefactos de software de monitoreo y autoconsciencia de un sistema IoT, a nivel de las capas Edge, Fog y Cloud. Por su parte, los artefactos generados por el middleware proveen el soporte necesario para que un sistema IoT pueda aprender y razonar en tiempo de ejecución mediante la parametrización de modelos, es decir, la actualización y calibración permanente de las métricas, metadatos y otros parámetros especificados en los modelos, de tal forma que estos modelos actualizados reflejen en todo momento el estado y posible evolución del sistema para la toma de decisiones.

3.4. Visor web de métricas de autoconsciencia. Aplicación web a nivel de front-end de la infraestructura de soporte que visualiza, de manera tabular y gráfica, las métricas, metadatos y recomendaciones resultantes de los procesos de autoconsciencia pre-reflexivos y reflexivos ejecutados por el sistema IoT.

4. ***Evaluación empírica de la solución tecnológica.*** Se han diseñado dos cuasi experimentos para evaluar la metodología Aware-IoT@Run.time y sus herramientas de soporte, los cuales siguen el proceso experimental de Wohlin et al. [34]. A su vez, estos cuasi experimentos utilizan modelos teóricos adaptados del MEM [33] que proporcionan un sustento científico sólido para comprobar el rendimiento (eficacia, eficiencia) del usuario al utilizar la solución Aware-IoT@Run.time, así como su percepción sobre la usabilidad, utilidad y posible adopción de la misma en proyectos futuros. No obstante, los procedimientos, modelos teóricos, instrumentos y materiales de evaluación diseñados en la tesis doctoral pueden ser tomados como base para diseñar nuevos experimentos que permitan evaluar otros enfoques metodológicos o herramientas de ingeniería de software para predecir su aceptación en la práctica.

5. Líneas de investigación futuras

A partir del desarrollo de la tesis doctoral han surgido importantes oportunidades de investigación que, si bien están fuera de los objetivos y alcance de esta tesis, tienen como finalidad mejorar y ampliar la solución metodológica Aware-IoT@Run.Time. A continuación, se resumen las principales líneas de trabajo que se plantean abordar en proyectos futuros:

- Continuar con las tareas de perfeccionamiento de la solución Aware-IoT@Run.Time, tendientes a optimizar el proceso metodológico y mejorar la usabilidad de las herramientas de soporte con base en los hallazgos y recomendaciones obtenidos de las evaluaciones.
- Completar la construcción de las funcionalidades y componentes de la infraestructura de soporte relacionados con los aspectos de autoconsciencia colectivos.
- Construir un Lenguaje de Modelado Específico de Dominio Visual (VDSML) basado en los metamodelos Aware-IoT y Monitor-IoT para diseñar gráficamente el dominio y las capacidades de autoconsciencia de un sistema IoT.
- La solución Aware-IoT@Run.Time se enmarca en el aprendizaje y razonamiento de los sistemas IoT mediante la parametrización de los modelos en tiempo de ejecución. Sin embargo, uno de los

desafíos a explorar a futuro es el aprendizaje y razonamiento de estos sistemas fundamentado en la estructura de los modelos. Para ello, se requiere de sistemas con capacidades de automodelado, los cuales deben estar habilitados para construir y mantener el modelado de sí mismo y su entorno durante el tiempo de ejecución [35]. En particular, debido a la falta de herramientas de ingeniería de software en este ámbito, las nuevas investigaciones deberán concentrarse en construir e integrar nuevos componentes a la infraestructura de soporte de Aware-IoT@Run.Time, encargados principalmente de: i) descubrir en tiempo real nuevas entidades IoT (físicas, digitales) y sus propiedades, o detectar cambios en estas; y, ii) actuar como modeladores no humanos, actualizando en tiempo de ejecución los modelos a nivel estructural (p. ej., agregar nuevas entidades IoT, propiedades, metadatos, flujos de datos, procesos de autoconsciencia, métricas, entre otros), de tal forma que reflejen los cambios que ocurren en el propio sistema y su entorno.

- Ampliar el alcance de la solución tecnológica para que adicional a los procesos de monitoreo y autoconsciencia, soporte la implementación de los procesos de actuación de un sistema IoT. De esta forma, el conocimiento cuantitativo y cualitativo sobre el estado y la posible evolución de un sistema IoT y su entorno, obtenido a partir de sus procesos de monitoreo y autoconsciencia, más allá de servir para llegar a recomendaciones, permitirá disparar automáticamente procesos de adaptación en tiempo real, otorgándole al sistema IoT propiedades tanto de autoconsciencia como autoadaptación. En este punto, la funcionalidad genérica del middleware de la infraestructura será valiosa para implementar los artefactos de software que soporten los procesos de actuación del sistema IoT, conforme a modelos de autoadaptación que deberán ser construidos.
- Realizar nuevos estudios de caso o experimentos en contexto industriales con la participación de profesionales experimentados; y, considerando escenarios de monitorización y autoconsciencia de mayor complejidad, incluso en otros subdominios de IoT.

6. Bibliografía

[1] *Internet of Things (IoT)—Reference Architecture, Standard ISO/IEC 30141:2018*. ISO, Aug. 2018. [Online]. Available: https://www.iso.org/standard/65695.html

[2] *Overview of the Internet of Things, Recommendation ITU-T Y.2060*. ITU-T Study Group 13, Jun. 2012. [Online]. Available: https://www.itu.int/rec/T-REC-Y.2060-201206-I

[3] R. Hassan, F. Qamar, M. K. Hasan, A. H. M. Aman, and A. S. Ahmed, "Internet of things and its applications: A comprehensive survey," *Symmetry*, vol. 12, no. 10, 2020.

[4] R. van der Meulen. *Gartner Says 8.4 Billion Connected 'Things' Will be in Use in 2017, up 31 Percent From 2016*. Gartner, Egham, U.K. Accessed: Feb. 7, 2017. [Online]. Available: https://www.gartner.com/en/newsroom/press-releases/2017-02-07-gartner-says-8-billion-connected-things-will-be-in-use-in-2017-up-31-percent-from-2016

[5] E. Estopace. *IDC Forecasts Connected IoT Devices to Generate 79.4 ZB of Data in 2025*. FutureIoT. Accessed: Jun. 22, 2019. [Online]. Available: https://futureiot.tech/idc-forecasts-connected-iot-devices-to-generate-79-4zb-of-data-in-2025

[6] D. Debnath and S. K. Chettri, "Internet of Things: Current Research, Challenges, Trends and Applications," in *First Global Conference on Artificial Intelligence and Applications (GCAIA)*, Singapore, 2021, pp. 679–694.

[7] D. Miorandi, S. Sicari, F. De Pellegrini, and I. Chlamtac, "Internet of Things: Vision, applications and research challenges," *Ad Hoc Networks*, vol. 10, no. 7, pp. 1497–1516, 2012.

[8] P. R. Lewis, "Self-aware computing systems: From psychology to engineering," in *Design, Automation & Test in Europe Conference & Exhibition (DATE)*, Lausanne, Switzerland, 2017, pp. 1044-1049.

[9] C. M. Sosa-Reyna, E. Tello-Leal, and D. Lara-Alabazares, "Methodology for the model-driven development of service oriented IoT applications," *Journal of Systems Architecture*, vol. 90, pp. 15–22, 2018.

[10] P. R. Lewis, M. Platzner, B. Rinner, J. Tørresen, and X. Yao, *Self-aware computing systems: An Engineering Approach*. Natural Computing Series. Cham, Switzerland: Springer, 2016.

[11] X. Chen, A. Li, X. Zeng, W. Guo, and G. Huang, "Runtime model based approach to IoT application development," *Frontiers of Computer Science*, vol. 9, no. 4, pp. 540–553, 2015.

[12] L. Esterle and R. Grosu, "Cyber-physical systems: challenge of the 21st century," *Elektrotechnik & Informationstechnik*, vol. 133, no. 7, pp. 299–303, 2016.
[13] P. Lalanda, S. Chollet, and C. Hamon, "Leveraging design and runtime architecture models to support self-awareness," in *Self-Aware Computing Systems*, S. Kounev, J. Kephart, A. Milenkoski, and X. Zhu, Eds. Cham, Switzerland: Springer, 2017, pp. 669–686.
[14] S. Kounev *et al.*, "The notion of self-aware computing," in *Self-Aware Computing Systems*, S. Kounev, J. Kephart, A. Milenkoski, and X. Zhu, Eds. Cham, Switzerland: Springer, 2017, pp. 3–16.
[15] P. R. Lewis *et al.*, "Towards a framework for the levels and aspects of self-aware computing systems," in *Self-Aware Computing Systems*, S. Kounev, J. Kephart, A. Milenkoski, and X. Zhu, Eds. Cham, Switzerland: Springer, 2017, pp. 51–85.
[16] J. Cámara *et al.*, "Self-aware computing systems: Related concepts and research areas," in *Self-Aware Computing Systems*, S. Kounev, J. Kephart, A. Milenkoski, and X. Zhu, Eds. Cham, Switzerland: Springer, 2017, pp. 17–49.
[17] A. Agarwal, J. Miller, J. Eastep, D. Wentziaff, and H. Kasture, "Self-aware computing," MIT, Tech. Rep. AFRL-RI-RS-TR-2009-161, 2009.
[18] A. Elhabbash, M. Salama, R. Bahsoon, and P. Tino, "Self-awareness in Software Engineering: A systematic literature review," *ACM Transactions on Autonomous and Adaptive Systems*, vol. 14, no. 2, pp. 1–42, 2019.
[19] T. Stahl and M. Volter, *Model-driven software development: Technology, Engineering, Management.* New York, NY, USA: John Wiley and Sons, 2006.
[20] J. Bézivin, "Model driven engineering: An emerging technical space," in *Generative and Transformational Techniques in Software Engineering (GTTSE). Lecture Notes in Computer Science*, vol. 4143, R. Lämmel, J. Saraiva, J. Visser, Eds. Berlin, Heidelberg: Springer, 2006, pp. 36–64.
[21] R. France and B. Rumpe, "Model-driven development of complex software: A research roadmap," in *Future of Software Engineering (FOSE)*, Minneapolis, MN, USA, 2007, pp. 37–54.
[22] M. Szvetits and U. Zdun, "Systematic literature review of the objectives, techniques, kinds, and architectures of models at runtime," *Software & Systems Modeling*, vol. 15, pp. 31–69, 2016.
[23] G. Blair, N. Bencomo, and R. France, "Models@run.time," *Computer*, vol. 42, no. 10, pp. 22–27, 2009.
[24] N. Bencomo, R. France, B. H. C. Cheng, and U. Aßmann, *Models@run.time: Foundations, applications, and roadmaps (vol. 8378)*. Cham, Switzerland: Springer, 2014.
[25] N. Bencomo, S. Götz, and H. Song, "Models@run.time: a guided tour of the state of the art and research challenges," *Software & Systems Modeling*, vol. 18, pp. 3049–3082, 2019.
[26] N. Bencomo, G. Blair, S. Götz, B. Morin, and B. Rumpe, "Report on the 7th international workshop on models@run.time," *ACM Software Engineering Notes*, vol. 38, no. 1, pp. 27–30, 2013.
[27] L. Erazo-Garzón, "Metodología Basada en Modelos en Tiempo de Ejecución para la Construcción y Operación de Sistemas Autoconscientes de Internet de las Cosas," in *XXV Ibero-American Conference on Software Engineering, Doctoral Symposium*, 2022, pp. 392–399.
[28] L. Erazo-Garzón, "Metodología basada en modelos en tiempo de ejecución para la construcción y operación de sistemas autoconscientes de Internet de las Cosas," Ph.D. dissertation, Facultad de Informática, Universidad de La Plata, La Plata, Argentina, 2024.
[29] F. García *et al.*, *Una Ontología de la Medición del Software*. Departamento de Informática, Universidad de Castilla-La Mancha, 2004.
[30] *Software & Systems Process Engineering Meta Model Specification Version 2.0*. Object Management Group-OMG. Apr. 2008. [Online]. Available: https://www.omg.org/spec/SPEM/2.0/PDF
[31] N. Mohan and J. Kangasharju, "Edge-Fog cloud: A distributed cloud for Internet of Things computations," in *2016 Cloudification of the Internet of Things (CIoT)*, Paris, France, *2016*, pp. 1-6.
[32] L. Erazo-Garzón, P. Cedillo, G. Rossi, and J. Moyano, "A Domain-Specific Language for Modeling IoT System Architectures That Support Monitoring," *IEEE Access*, vol. 10, pp. 61639–61665, 2022.
[33] D. L. Moody, G. Sindre, T. Brasethvik, and A. Solvberg, "Evaluating the quality of information models: empirical testing of a conceptual model quality framework," in *25th Int. Conf. on Software Engineering*, Portland, OR, USA, 2003, pp. 295-305.
[34] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén, *Experimentation in Software Engineering*. Berlin, Germany: Springer, 2012.
[35] K. L. Bellman et al., "Self-modeling and self-awareness," in Self-Aware Computing Systems, S. Kounev, J. Kephart, A. Milenkoski, and X. Zhu, Eds. Cham, Switzerland: Springer, 2017, pp. 279–304.

Métodos computacionales avanzados para evaluar la correlación temporal entre apellidos y ancestría biológica en Argentina.

Autor: Arturo Leonardo Morales

Directores: Director: Dr. Claudio Delrieux - Co-Directora: Dra. Virginia Ramallo

Universidad: Universidad Nacional de la Patagonia San Juan Bosco

Fecha de exposición: 30 de julio de 2024

Resumen

Motivación

Esta tesis aborda el estudio de la estructura poblacional argentina mediante el análisis de apellidos complementado con otras fuentes de datos públicos no estructurados. Su propósito es examinar el impacto del procesamiento avanzado y el modelado computacional de apellidos en la generación de indicadores precisos sobre ancestría biológica, dinámica poblacional y estructura demográfica.

El uso de apellidos ofrece ventajas metodológicas clave:

- Se trata de un recurso accesible y de bajo costo para el análisis poblacional.
- Permiten una cobertura amplia en términos geográficos y temporales.
- Facilitan la integración con otras fuentes de información, como registros censales, encuestas e informes técnicos y datos sanitarios.

A partir de estas características, nuestra investigación explora la correlación entre apellidos, estructura genética, salud pública y migración, proporcionando herramientas analíticas innovadoras para la comprensión de la dinámica poblacional en Argentina.

Aporte a la disciplina

En nuestro trabajo aplicamos métodos computacionales avanzados que integran Big Data, Minería de Datos, Aprendizaje Automático y Sistemas de Información Geográfica (SIG) para mejorar el análisis de la diversidad poblacional. Entre sus principales contribuciones se destacan:

- Diseño e implementación de algoritmos eficientes para el etiquetado de apellidos según su origen geográfico y lingüístico, en conjunto con su visualización y análisis de dispersión espacial.
- Desarrollo de modelos computacionales que correlacionan apellidos con datos de salud pública y patrones migratorios.

- Creación de herramientas interactivas para la exploración y visualización de datos, optimizando el análisis epidemiológico y demográfico, entre las que se destaca la aplicación web prototipo denominada *Bulsarapp* (Fig 1).
- Diseño y creación de biblioteca de funciones para el lenguaje de programación *Python*, denominado "*isonymic*", para la aplicación eficiente y reproducible del método isonímico (Fig 2). Este paquete facilita la inferencia de niveles de endogamia, consanguinidad y movilidad poblacional en diferentes regiones del país.
- Recopilación, validación exhaustiva y publicación de los conjuntos de datos sobre migración y aislamiento que fueron derivados del análisis de apellidos. Este proceso constituye un paso fundamental para consolidar esta nueva fuente de datos como un insumo clave para los estudios poblacionales. Su cobertura, tanto retrospectiva como prospectiva, permitirá a los investigadores comprender con mayor precisión los patrones demográficos y sociales que influyen en la población.

Las herramientas obtenidas y las metodologías optimizadas pueden ampliarse, escalarse y adaptarse a otros países y contextos, constituyendo un marco flexible para el análisis de poblaciones en distintos niveles administrativos.

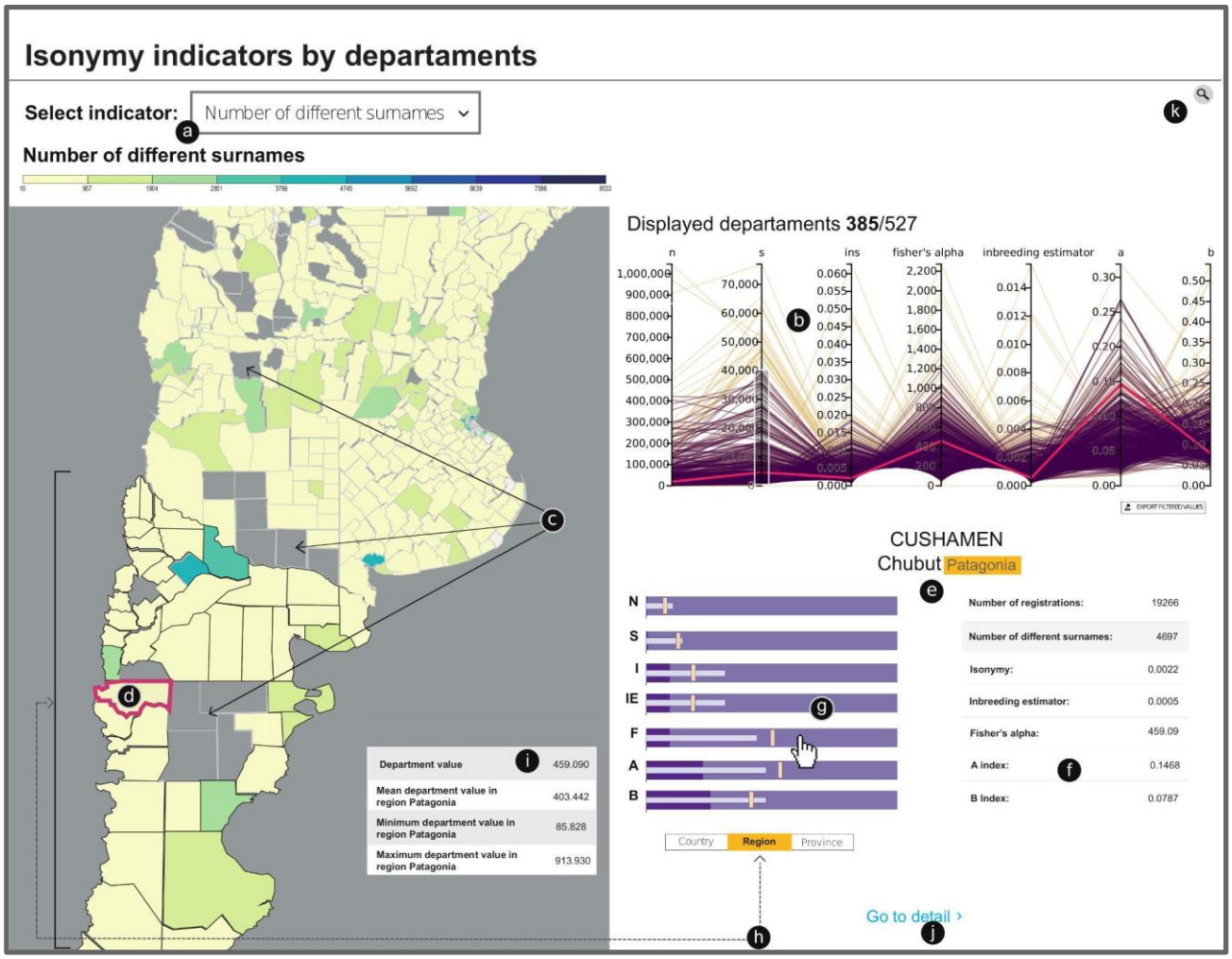


Fig 1: Una vista de la aplicación web Bulsarapp. La herramienta permite interactuar con el mapa mientras con otras interacciones que permiten: a) Seleccionar cualquiera de los índices isonímicos. b) Filtrar rangos en un gráfico de coordenadas paralelas. c) Aplicar filtros en los ejes del gráfico de coordenadas paralelas interactivas, resaltando la línea de la unidad territorial seleccionada. d) Observar la vecindad de la unidad elegida a nivel de país, región o provincia. Se destaca la vecindad con bordes oscuros en el mapa. e) Evaluar los valores de isonimia, totales (f) y contextualizados (g). i) Al pasar el cursor sobre las barras del gráfico de viñetas se muestra un tooltip informativo y de resumen. h) Cambiar la contextualización regional de una unidad territorial. j) Ir a los detalles. k) Utilizar una herramienta de búsqueda en la esquina superior derecha para localizar departamentos o provincias por nombre.

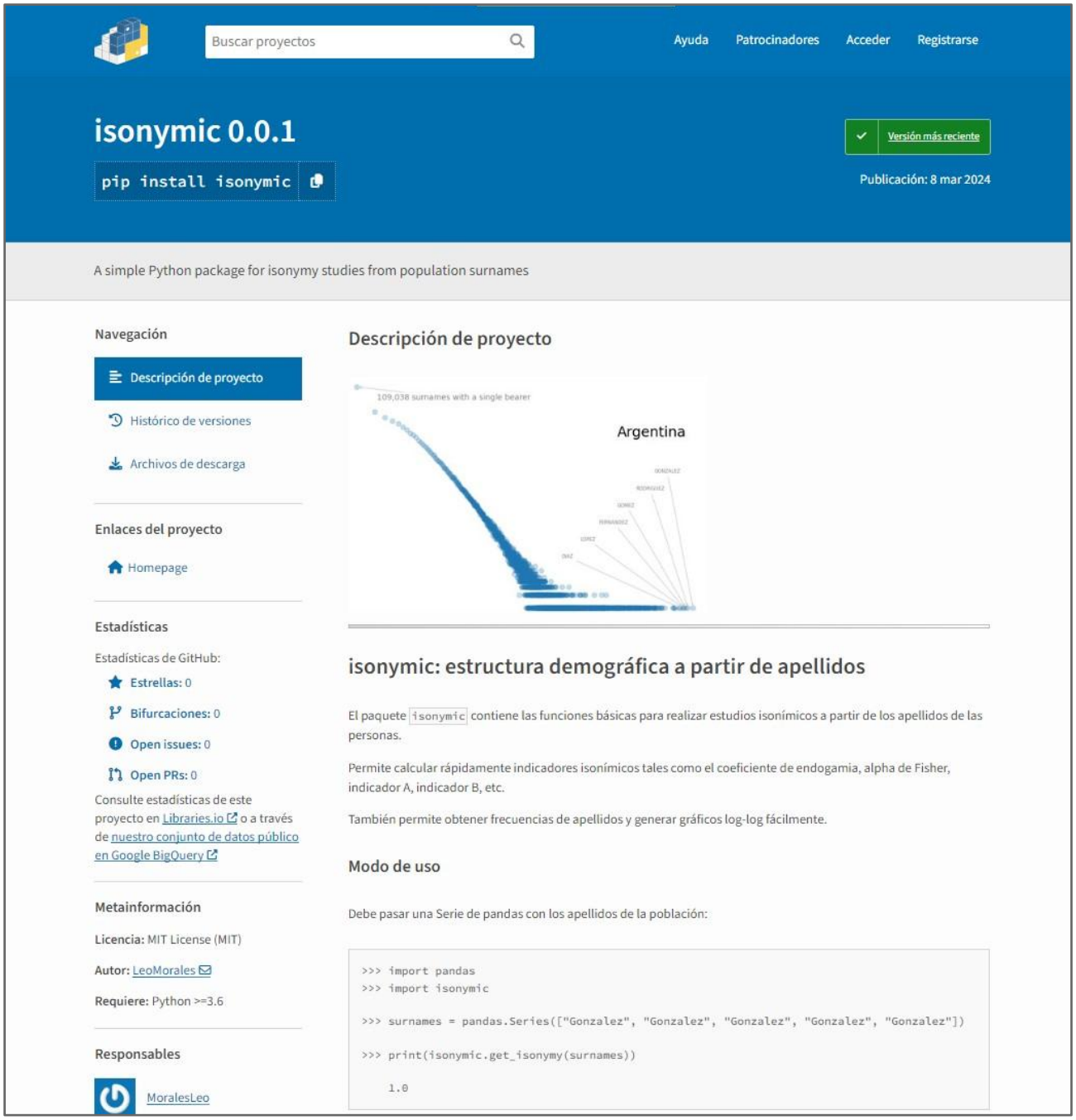


Fig 2: Paquete isonímico publicado en el portal PyPI, repositorio oficial de software para el lenguaje de programación *Python*.

Metodología y casos de estudio

El estudio se basa en fuentes de datos abiertas y preexistentes, incluyendo registros censales, padrones electorales y bases estadísticas de salud pública. En una primera fase, la investigación emplea apellidos como proxy de la estructura poblacional, desarrollando dos estudios clave:

- 1. Migración y salud: Se estudia la relación entre aislamiento poblacional y enfermedades neurodegenerativas, enfocándose en la comunidad de los Alemanes del Volga y la incidencia del Alzheimer en Argentina.
- 2. Migración interna en Argentina: Se analizan patrones espaciales de movilidad en un período de 20 años, utilizando datos de los padrones electorales de 2001, 2015 y 2021.

Posteriormente, nuestra tesis amplía su alcance hacia el análisis epidemiológico, optimizando y extendiendo los métodos desarrollados para abordar nuevas preguntas en salud pública. Se presentan tres estudios adicionales:

- 1. Epidemiología de enfermedades poco frecuentes: Se investigó la relación entre estructura poblacional y distribución de enfermedades genéticas raras, utilizando las metodologías computacionales ya desarrolladas para estudios con apellidos pero adaptadas a bases de datos de salud pública y registros especializados como *Orphanet*. Nuestra estrategia fue abordar el tema a través de registros de fallecimientos detallados para el período de años comprendido entre el 1991 y el 2017.

  Si bien se identificaron agrupamientos espaciales asociados a estas patologías, sus múltiples dimensiones, producto del trazado de 5885 enfermedades mapeadas a códigos CIE-10, hacen evidente la necesidad de profundizar el análisis. Dicha línea podría enfocarse en posibles correlaciones con patrones migratorios históricos y contemporáneos, así como con eventos significativos en el desarrollo social de nuestro país.

- 2. Epidemiología de muertes fetales: Se analizó la variación espacial y temporal de las tasas de mortalidad fetal en Argentina, utilizando los modelos geoespaciales que se aplicaron a los dominios anteriores pero incorporando otros datos sanitarios.

  Se evaluó el impacto de factores socioeconómicos y demográficos, utilizando herramientas de análisis estadístico y minería de datos para identificar patrones de riesgo.

- 3. Análisis espacial de la bronquiolitis a nivel urbano: En este estudio investigamos la incidencia, a una escala más fina como es la del entorno urbano, de una enfermedad puntual, la bronquiolitis, y con un enfoque en la correlación entre su distribución y la de variables socioeconómicas y ambientales. Se analizaron específicamente los casos de internación en el principal y único hospital público de Puerto Madryn, una ciudad que ha experimentado un crecimiento significativo, al sur de la Argentina. Se aplicaron modelos de georreferenciación para detectar áreas de mayor vulnerabilidad y evaluar el impacto del hacinamiento en la propagación de la enfermedad.

Conclusiones y contribución final

Los resultados de esta tesis demuestran el potencial de integrar bases de datos preexistentes con métodos computacionales avanzados para profundizar la comprensión de la dinámica poblacional. La propuesta metodológica resalta varios aspectos clave:

- Es de bajo costo y replicable en el tiempo, ya que utiliza fuentes de datos ya disponibles sin requerir recolección específica.
- Es complementaria a estudios tradicionales, aportando nuevas perspectivas para el análisis poblacional y epidemiológico.
- Y resulta escalable a nivel internacional, permitiendo su adaptación a otros países y contextos de investigación.

Asimismo, se desarrollaron y evaluaron, en colaboración con expertos del área, un conjunto de herramientas especializadas para la generación de reportes, visualizaciones interactivas y análisis de patrones espaciales, facilitando la toma de decisiones en salud pública y demografía. Estas herramientas, junto con los datos generados a partir de los análisis derivados de su aplicación y el aprovechamiento de nuevas fuentes de información, pueden servir como insumo para las avanzadas herramientas de inteligencia artificial contemporáneas, con el propósito de fortalecer el análisis predictivo y optimizar la formulación de políticas basadas en evidencia.

Líneas de investigación futuras

Este trabajo abre nuevas posibilidades de estudio en diversas áreas:

- Extensión del análisis a otros países de América Latina para identificar patrones migratorios y estructuras poblacionales transnacionales.
- Incorporación de técnicas avanzadas de Inteligencia Artificial para mejorar la predicción de tendencias poblacionales y epidemiológicas.
- Aplicación del modelo a estudios genéticos y de salud pública, evaluando la distribución de enfermedades hereditarias en distintas poblaciones. En este contexto, para nuestro grupo de investigación, el estudio de caso EPOF abordado en esta tesis representó un primer paso hacia un análisis en desarrollo, orientado a desentrañar la complejidad del fenómeno estudiado y a explorar posibles correlaciones con patrones migratorios históricos y contemporáneos, así como con movimientos clave vinculados al desarrollo social de nuestro país.

Esta tesis representa un avance significativo en la intersección entre ciencias de la computación, genética de poblaciones, epidemiología y demografía, consolidando nuevas herramientas y procedimientos para el análisis de datos a gran escala.

Minería de Textos para la extracción de conocimiento en actividades educativas con información proveniente de redes sociales. Una estrategia didáctica

Resumen de la Tesis presentada para obtener el grado de Doctor en Ciencias Informáticas

Posgrado, Facultad de Informática, Universidad Nacional de La Plata (UNLP)

Fecha de defensa: 12 de abril de 2024

Autora: Ana Lucía Pérez Suasnavas, alperezs@uce.edu.ec[1]

Directora: Karina Cela, klcela@espe.edu.ec[2]

Codirector: Waldo Hasperué, whasperue@lidi.info.unlp.edu.ar[3]

[1] Universidad Central del Ecuador, Facultad de Ingeniería y Ciencias Aplicadas

[2] Universidad de las Fuerzas Armadas ESPE, Departamento de Computación

[3] Universidad Nacional de La Plata, Facultad de Informática, Instituto de Investigación en Informática (III-LIDI). Comisión de Investigaciones Científicas (CIC)

1. Resumen

La tesis doctoral se desarrolló en el campo de las tecnologías aplicadas a la educación, de forma particular, mediante el uso de técnicas de Minería de Datos y la integración de redes sociales.

En el área de la educación, se ha evidenciado el uso de distintas técnicas de Minería de Datos, con diferentes objetivos. Sin embargo, pocos estudios refieren al uso de prácticas pedagógicas que, integren el uso de técnicas de Minería de Datos y la red social Twitter.

El objetivo de la presente tesis doctoral, consiste en proponer una estrategia didáctica que integre técnicas de Minería de Datos para predecir las dificultades que experimentan los estudiantes universitarios en una sesión de clase, mediante la extracción de datos provenientes de la red social Twitter; con el propósito de que, el docente cuente con una herramienta que le permita brindar ayuda oportuna a los estudiantes, en los temas que demandan mayor atención en la enseñanza de la asignatura de Programación 1, en la Universidad Central del Ecuador, localizada en la ciudad de Quito; donde, uno de los principales problemas que enfrenta esta institución, es el volumen de estudiantes que existen en cada curso; limitando la posibilidad del docente en mantener un seguimiento efectivo del desempeño académico de los estudiantes.

2. Palabras clave

Minería de Textos; Twitter; Estrategia Didáctica; Redes Neuronales Artificiales; Algoritmos Supervisados; Procesamiento de Lenguaje Natural; APP Issues; JiTTwT.

3. Motivación

Con la creciente demanda y uso de las redes sociales, así como el manejo de diferentes medios digitales, el volumen de la información en la Web ha incrementado significativamente (Fan y Bifet, 2013); por lo que, el almacenamiento, la extracción y el análisis de estos grandes conjuntos de datos requieren de nuevas herramientas para procesar la información, debido a que son datos provenientes de diversas fuentes, son datos heterogéneos y no cuentan con una estructura específica (Matas Terrón et al., 2020; Tiwari y Kumar, 2020; Zong et al., 2021).

No obstante, estos conjuntos de datos heterogéneos, pueden ser aplicados en diferentes áreas, como las finanzas, marketing, salud, industria, educación, entre otras (Aggarwal, 2011). Para el aprovechamiento de estos datos, existen múltiples técnicas con propósitos diferentes como: análisis estadístico de redes, detección de comunidades, clasificación de la información, preservación la privacidad de la información, análisis de tendencia, entre otros.

En este contexto, para el uso efectivo de las redes sociales, es necesario que las técnicas actuales, permitan la extracción, depuración y análisis de la información para alcanzar el descubrimiento de la información, objetivo principal del KDD «Knowledge Discovery in Databases»; el cual, luego de aplicar sus diferentes etapas permitirá interpretar los resultados para la toma de decisiones futuras (Hernández Orallo et al., 2004; Kantardzic, 2020; Tan et al., 2014)..

Una de las áreas en las que se han aplicado las redes sociales es la educación; donde se ha evidenciado que el propósito general es mejorar los procesos de enseñanza y aprendizaje en los diferentes niveles educativos (Anisimova et al., 2019; Chugh y Ruhi, 2018; Lemay et al., 2019; Tejedor et al., 2021). Entre las redes sociales que más destacan en los últimos años son Facebook y Twitter (Froment et al., 2022; Marín-Díaz y Cabero-Almenara, 2019); donde Twitter predomina en la educación superior (Fernández-Ferrer y Cano, 2016). Aunque los estudios demuestran el principal aporte de Twitter en niveles superiores de educación, no ha sido posible determinar estrategias didácticas que permitan estandarizar su aplicación (Marín-Díaz y Cabero-Almenara, 2019; Tejedor et al., 2021); que permitan fomentar una cultura digital entre los docentes, estudiantes y la institución; frente a un gran volumen de información que derivan las redes sociales (Salinas Ibáñez y Marín Juarros, 2019).

Frente a estos escenarios, este trabajo plantea la necesidad de desarrollar una estrategia que permita integrar aspectos pedagógicos, medios tecnológicos y técnicas de Minería de Datos para el aprovechamiento de los datos provenientes de las redes sociales; para la inserción en el contexto educativo, de forma que, se derive una propuesta didáctica que logre alcanzar una mejora de los procesos de enseñanza y aprendizaje.

4. Aporte de la tesis a la disciplina

Mediante el desarrollo y aplicación de la estrategia didáctica, se pretende alcanzar diferentes logros que beneficien a los actores de la educación. Al considerarse una propuesta, los actores que intervienen en este estudio son: el docente y los estudiantes. A continuación, se procede a describir brevemente los principales beneficios que obtendrían cada uno:

a) Beneficios para el Docente

- Conseguir la motivación en el aula.
- Usar efectivamente las redes sociales en el aula.
- Conseguir que el porcentaje de estudiantes reprobados en la asignatura disminuya.
- Anticipar las dificultades que los estudiantes puedan expresar, y brindar ayuda oportuna en los temas que demanden más atención.

b) Beneficio para los Estudiantes

- Población beneficiada: estudiantes de los primeros niveles de formación, esto es 2° a 3° nivel, de carreras de Ingeniería, donde se imparta la cátedra de Programación 1.
- Conocer los conceptos generales de Programación 1.
- Desarrollar diferentes habilidades cognitivas (análisis, razonamiento lógico, abstracción, creatividad, etc.); y habilidades de trabajo en equipo; que pueden ser aplicadas a cualquier asignatura.
- Participar activamente dentro y fuera del aula.
- Focalizar el aprendizaje basado en el estudiante.
- Usar las redes sociales con fines académicos.
- Mejorar el rendimiento académico y aprobar la asignatura.

Las principales contribuciones que derivó la tesis doctoral, se resumen en los siguientes puntos:

- Una revisión sistemática de la literatura actualizada al año 2022 y con datos recopilados desde hace 15 años atrás.
- Cuestionario “Social Network & Education” para identificar la tendencia de uso de las redes sociales y aspectos sobre la participación, motivación y asistencia a las clases, validado por expertos y medido con Alfa de Cronbach, por lo que se considera un instrumento fiable y válido.
- Una Estrategia Didáctica denominada JiTTwT para fomentar la participación estudiantil durante las sesiones de clases presenciales; que integra prácticas pedagógicas, la red social Twitter y técnicas de Minería de Datos; y que puede ser aplicada a cualquier asignatura de una carrera universitaria.
- Diagrama de Flujo de la Estrategia Didáctica JiTTwT que permite condensar el proceso y las actividades que incluye la estrategia (Figura 1).
- Cuestionario “Actitud hacia el aprendizaje de Programación y Autopercepción al Pensamiento Crítico”; para identificar la postura del estudiante en la asignatura, y la habilidad del pensamiento crítico; el cual fue validado por expertos y medido con Alfa de Cronbach, por lo que se considera un instrumento fiable y válido.
- Un Corpus Especializado Ad-hoc en la asignatura de Programación 1, escrito en idioma español.
- Un Conjunto de Datos «DataSet» que contiene un registro de las dificultades estudiantiles en la asignatura de Programación 1, clasificadas por tipo de dificultad, escrito en idioma español.
- Una de las funcionalidades de la aplicación informática «API Issues» para la extracción de los datos provenientes de Twitter de una cuenta específica.

- La segunda funcionalidad de la «API Issues» para el preprocesamiento de los datos, de forma que se cuente con un «DataSet» depurado. Esta funcionalidad de la aplicación, puede ser utilizada en cualquier área que requiera el preprocesamiento de los datos.
- La tercera funcionalidad de la «API Issues» permite la construcción y evaluación del modelo predictivo.
- La cuarta funcionalidad de la «API Issues» predice y visualiza las dificultades académicas reportadas por los estudiantes, y que requieren mayor atención por parte del docente.
- Un Flujo de Trabajo que integra técnicas de Minería de Textos, Procesamiento de Lenguaje Natural, Aprendizaje Profundo y estrategias didácticas para predecir las dificultades estudiantiles, no basadas únicamente en aspectos académicos, sino también personales e institucionales (Figura 2).

5. Posibles líneas de investigación futuras

- Aplicar el instrumento "ENCUESTA PARA MEDIR LA ACTITUD HACIA EL APRENDIZAJE DE PROGRAMACIÓN Y LA AUTOPERCEPCIÓN DEL PENSAMIENTO CRÍTICO" a otros grupos de estudio, de forma que puedan compararse los resultados de esta investigación con nuevos hallazgos.
- Analizar los factores que inciden en la actitud hacia el aprendizaje de la asignatura de Programación 1 en la educación superior, para determinar la influencia de la aplicación de la Estrategia Educativa JiTTwT.
- Analizar las causas por las que, el rendimiento de los estudiantes disminuye en el segundo Hemi, con relación al primer Hemi; sin considerar la pertenencia al grupo de control o al grupo experimental.
- JiTTwT al ser una propuesta de estrategia didáctica, requiere ser validada con nuevos casos de estudio; puesto que, como bien lo señalan Forés Miravalles y Bueno i Torrens (2019), se necesitan datos y tiempo para afianzarla; lo básico es recolectar evidencias, tener parámetros, voces, experiencias; y en el futuro se pueda consolidar como una metodología.
- Utilizar la Estrategia Didáctica JiTTwT en otras instituciones de educación superior, con el objetivo de verificar la aplicabilidad y observar las respuestas de los estudiantes en otros entornos de trabajo, ya sea a nivel público o privado; y observar las posibles diferencias o semejanzas encontradas.
- Identificar los motivos del alto porcentaje de reprobación de la asignatura de Programación 1 en la Carrera de Ingeniería Civil, así como proponer alternativas para minimizar este problema.
- Analizar otros factores que inciden en la aplicación de la Estrategia Didáctica JiTTwT, como son: estudiantes que repiten la asignatura y el rendimiento académico.
- Examinar el estilo de aprendizaje de los estudiantes, antes de la aplicación de la Estrategia Didáctica JiTTwT, de forma que se puede llevar un control sobre la incidencia o no de este factor.
- Aunque el porcentaje de participación estudiantil incrementó, al igual que el valor de la media para el grupo experimental, se puede evidenciar que la desviación estándar también incrementó, esto permite inferir que existe mayor dispersión de los datos respecto del registro

de participaciones para este grupo, este resultado requiere de un análisis posterior de las causas asociadas.

- Analizar las causas de la existencia de un grupo experimental y de control, que, aunque no participaron y no mejoran el rendimiento, igualmente aprueban la asignatura de Programación 1.
- Como condición de mejora, sería el análisis de la aplicación de la Estrategia didáctica JiTTwT y su incidencia en otros procesos de enseñanza y aprendizaje como la evaluación y la retención.
- En el análisis antes y durante la pandemia, se consideraron los tipos de dificultades con mayor porcentaje de ocurrencias, estos son: el tipo D003 y el tipo D005; sin embargo, conviene analizar los otros tipos de dificultades y evaluar su evolución entre períodos.
- Combinar técnicas de agrupamiento (no incluye visualización de la información), con las de clasificación, de forma que, se evite la tarea de etiquetar manualmente los registros de las dificultades de los estudiantes. Además, el agrupamiento permitiría crear taxonomías que ayudarían a crear categorías y subcategorías de temas afines.
- Aplicar otros algoritmos, como «Naïve Bayes», SVM o Árboles de decisión, que permita comparar los resultados y determinar el más efectivo para el caso de estudio, en función del cálculo de la exactitud del modelo predictivo.
- Utilizar las dos primeras funcionalidades de la API Issues en otros campos de estudio, para extraer y depurar de los datos, de forma que se puede conformar un nuevo «DataSet» de acuerdo con el objetivo que se persiga.
- Proponer herramientas o estrategias educativas, que permitan minimizar las dificultades encontradas en la última etapa del proceso del KDD, respecto de la enseñanza de la asignatura de Programación 1.
- Aplicar la herramienta API Issues para predecir el aprendizaje de los estudiantes; considerando que, en el presente trabajo no fueron considerados los comentarios sin dificultad (SDIF).

6. Referencias

Aggarwal, C. C. (2011). An introduction to social network data analytics. En C. C. Aggarwal (Ed.), *Social network data analytics* (pp. 1-15). Springer. https://doi.org/10.1007/978-1-4419-8462-3_1

Anisimova, O., Vasylenko, V., y Fedushko, S. (2019). Social networks as a tool for a higher education institution image creation. *arXiv preprint arXiv:1909.01678*. https://doi.org/10.48550/arXiv.1909.01678

Chugh, R., y Ruhi, U. (2018). Social media in higher education: A literature review of Facebook. *Education and Information Technologies*, *23*(2), 605-616. https://doi.org/10.1007/s10639-017-9621-2

Fan, W., y Bifet, A. (2013). Mining big data: Current status, and forecast to the future. *ACM SIGKDD explorations newsletter*, *14*(2), 1-5. https://doi.org/10.1145/2481244.2481246

Fernández-Ferrer, M., y Cano, E. (2016). The influence of the internet for pedagogical innovation: Using twitter to promote online collaborative learning. *International Journal of Educational Technology in Higher Education*, *13*(1), 22. https://doi.org/10.1186/s41239-016-0021-2

Forés Miravalles, A., y Bueno i Torrens, D. (2019, febrero 20). Hay una fórmula para saber si una metodología educativa tiene solvencia científica. *The Conversation*. https://theconversation.com/hay-una-formula-para-saber-si-una-metodologia-educativa-tiene-solvencia-cientifica-111969

Froment, F., García-González, A.-J., y Cabero-Almenara, J. (2022). Relación de la red social Twitter con la credibilidad docente y la motivación del alumnado universitario. *Revista Comunicar*, *30*(71), 131-142. https://doi.org/10.3916/C71-2022-10

Hernández Orallo, J., Ferri Ramirez, C., y Ramirez Quintana, M. J. (2004). *Introducción a la Minería de Datos* (Primera). Pearson Educación. http://bit.ly/2FGtqP1

Kantardzic, M. (2020). *Data mining: Concepts, models, methods, and algorithms* (Third Edition). John Wiley & Sons.

Lemay, D. J., Basnet, R. B., Doleck, T., y Bazelais, P. (2019). Social network analysis of twitter use during the AERA 2017 annual conference. *Education and Information Technologies*, *24*(1), 459-470. https://doi.org/10.1007/s10639-018-9787-2

Marín-Díaz, V., y Cabero-Almenara, J. (2019). Las redes sociales en educación: Desde la innovación a la investigación educativa. *Revista Iberoamericana de Educación a Distancia*, *22*(2), 25-33. https://doi.org/10.5944/ried.22.2.24248

Matas Terrón, A., Leiva Olivencia, J. J., y Franco Caballero, P. D. (2020). Big Data Irruption in Education. *Píxel-BIT Revista de Medios y Educación*, *57*, 59-90. https://doi.org/10.12795/pixelbit.2020.i57.02

Salinas Ibáñez, J. M., y Marín Juarros, V. I. (2019). Metasíntesis cualitativa sobre colaboración científica e identidad digital académica en redes sociales. *RIED. Revista Iberoamericana De Educación a Distancia*, *22*(2), 97-117. http://dx.doi.org/10.5944/ried.22.2.23238

Tan, P.-N., Steinbach, M., y Kumar, V. (2014). *Introduction to Data Mining*. Pearson.

Tejedor, S., Coromina, Ó., y Pla-Campas, G. (2021). Microblogging en escenarios curriculares universitarios: El uso de Twitter más allá del encargo docente. *Revista Electrónica de Investigación Educativa*, *23*, 1-13. https://doi.org/10.24320/redie.2021.23.e20.3565

Tiwari, D., y Kumar, M. (2020). Social media data mining techniques: A survey. En M. Tuba (Ed.), *Information and Communication Technology for Sustainable Development* (pp. 183-194). Springer. https://doi.org/10.1007/978-981-13-7166-0_18

Zong, C., Xia, R., y Zhang, J. (2021). *Text Data Mining*. Tsinghua University Press. https://doi.org/10.1007/978-981-16-0100-2_1

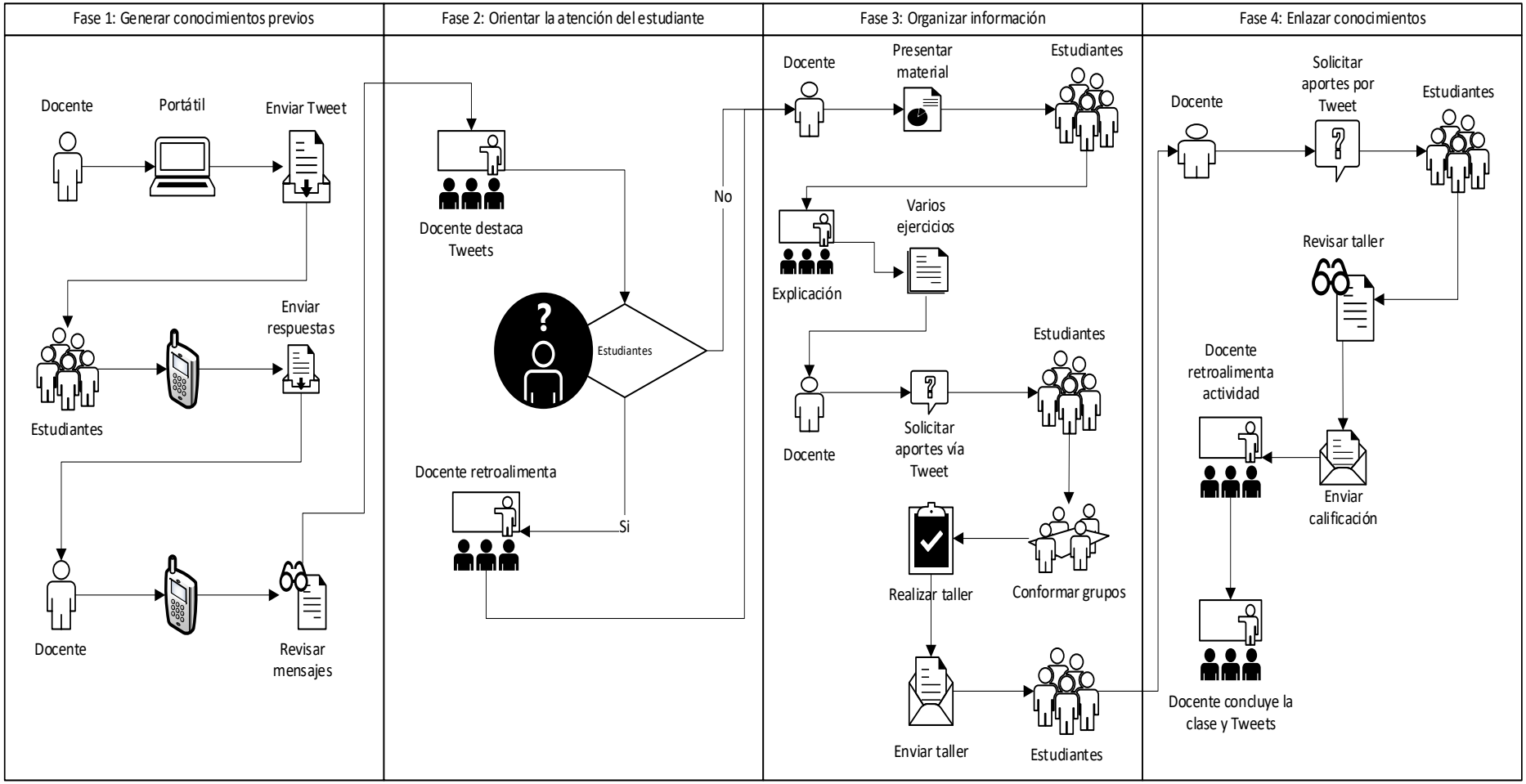


Figura 1 - Diagrama de Flujo de la Estrategia Didáctica JiTTwT

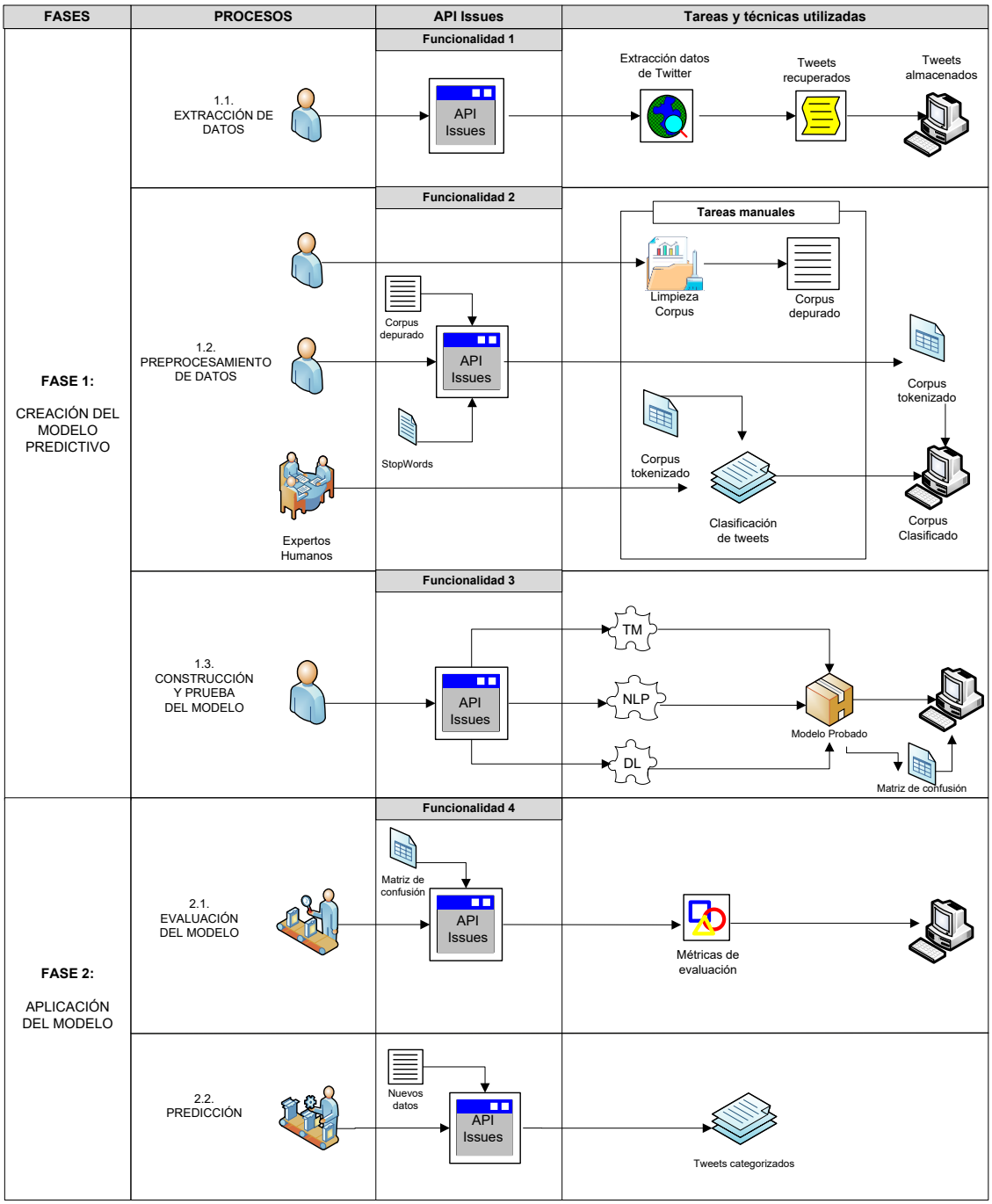


Figura 2 - Flujo de trabajo del modelo predictivo

Modelo de proceso para la construcción y mantenimiento de muestras de proyectos software para estudios en la Ingeniería del Software mpírica

Autor: Juan Andrés Carruthers

Director: Emanuel Agustín Irrazábal

Codirector: Jorge Andrés Diaz-Pace

Fecha de exposición: 4 de diciembre del 2024

Universidad Nacional del Nordeste

Resumen

Las muestras de proyectos software son una de las principales fuentes de información utilizadas en la Ingeniería del Software Empírica. En este sentido, el desarrollo de software de fuente abierta y su difusión a través de plataformas para compartir código han facilitado el acceso a millones de proyectos, sin embargo, la falta de controles en la creación de proyectos implica que algunos pueden contener información irrelevante. Otras fuentes son las muestras compartidas por la comunidad científica, aunque, en ocasiones, se conforman siguiendo procedimientos arbitrarios, con escasa documentación y cuestionables desde el punto de vista metodológico.

Para abordar estos problemas se propuso como solución la implementación del modelo de proceso SUM4SOFT para la construcción, actualización y curaduría de colecciones de proyectos software. El modelo se probó en un estudio de Ingeniería del Software y se refinó por medio de un estudio de caso. La aplicación del modelo de proceso resultó en un instrumento de recolección de datos y colecciones de proyectos software que se validaron en un estudio de minería de repositorios. Esto demostró que tanto SUM4SOFT como las herramientas implementadas constituyen soluciones viables para generar resultados confiables en la Ingeniería del Software Empírica.

Palabras clave

Muestra de proyectos de software, Actualización de muestras, Software de fuente abierta, Ingeniería del Software Empírica.

Motivación

La Ingeniería del Software Empírica (ISE) es una disciplina enfocada en recolectar y analizar datos obtenidos del proceso de desarrollo y mantenimiento del software para evaluar herramientas, prácticas o procedimientos que permitan, por ejemplo: corregir errores, detectar vulnerabilidades, estimar el tiempo de desarrollo, con el objetivo de mejorar la calidad del software integrando los hallazgos obtenidos como evidencia científica [1].

En particular, la ISE incluye métodos de investigación donde se esperan resultados generalizables y replicables (por ejemplo, experimentos, encuestas, estudios de minería de repositorios de software) [2]. La generalización implica que los hallazgos de una investigación se pueden aplicar a todos los sujetos de la población objetivo y para ello se requieren muestras representativas [3]. En cambio, la replicación consiste en validar los resultados de un estudio replicando una investigación bajo las mismas condiciones produciendo resultados consistentes con los originales [4]. Esto implica que los investigadores deben aplicar una estrategia de muestreo adecuada para la recolección de datos y registrar de manera completa y precisa el protocolo del estudio.

Una de las fuentes de datos empleadas para conducir estudios empíricos es el código fuente del proyecto software [5]. En este sentido, plataformas para compartir código de fuente abierta como GitHub, Gitlab o Bitbucket ofrecen acceso a millones de proyectos software para el desarrollo de estudios empíricos, los cuales deben ser seleccionados de manera cuidadosa porque muchos de ellos no son aptos para la investigación en ISE [6, 7]. Dado el esfuerzo que significa la recolección de datos, algunos estudios comparten a la comunidad científica las colecciones de proyectos software utilizadas en la investigación.

No obstante, estudios secundarios [8, 9] revelaron ciertas prácticas encontradas en la literatura que comprometen la validez de los resultados obtenidos a partir de estas colecciones. Por un lado, problemas metodológicos en la conformación de las muestras: número reducido de proyectos, sin criterios explícitos de evaluación de calidad y técnicas de muestreo no probabilísticas. Por otro lado, cuestiones sobre la difusión de los instrumentos de la investigación, como instrucciones incompletas, acceso limitado a los conjuntos de datos, ausencia de paquetes de replicación, entre otras.

Para abordar esta necesidad se propuso la construcción del modelo de proceso SUM4SOFT. El modelo detalla los pasos y reglas para construir colecciones representativas, y además ofrece plantillas para estandarizar la generación de los instrumentos de recolección de datos facilitando la replicación de los estudios. Para diseñar, construir y validar SUM4SOFT se empleó la metodología ciencia de diseño [10] compuesta de cinco actividades: explicar el problema, definir objetivo y requisitos, diseñar y desarrollar el artefacto, demostrar el artefacto y evaluar el artefacto. La Fig. 1 muestra las tareas planteadas para alcanzar el objetivo de la tesis en el marco de ciencia de diseño, donde se destacan cinco estudios empíricos: dos mapeos sistemáticos de la literatura, un estudio longitudinal, un estudio de minería de repositorios de software y un estudio de caso. Los objetivos de la investigación, los resultados preliminares y la planificación se presentaron en el simposio doctoral del Congreso Iberoamericano en Ingeniería del Software [11].

Colecciones en la ISE

Para conocer el estado de arte de la construcción de colecciones orientadas a estudios en ISE se condujeron dos mapeos sistemáticos [12]. Inicialmente, se analizaron estudios primarios que emplearon colecciones de proyectos software publicados en revistas y congresos especializados en ISE [13]. Durante la revisión, se detectó que el 28% de los artículos emplearon colecciones de referencia. Por ello, se decidió ampliar la revisión en un segundo mapeo sistemático enfocado en estas colecciones consumidas de manera frecuente por la comunidad científica para complementar los hallazgos iniciales [14]. En ambos casos se estudiaron los criterios de selección, la caracterización de los proyectos software, los metadatos extraídos y el uso de herramientas.

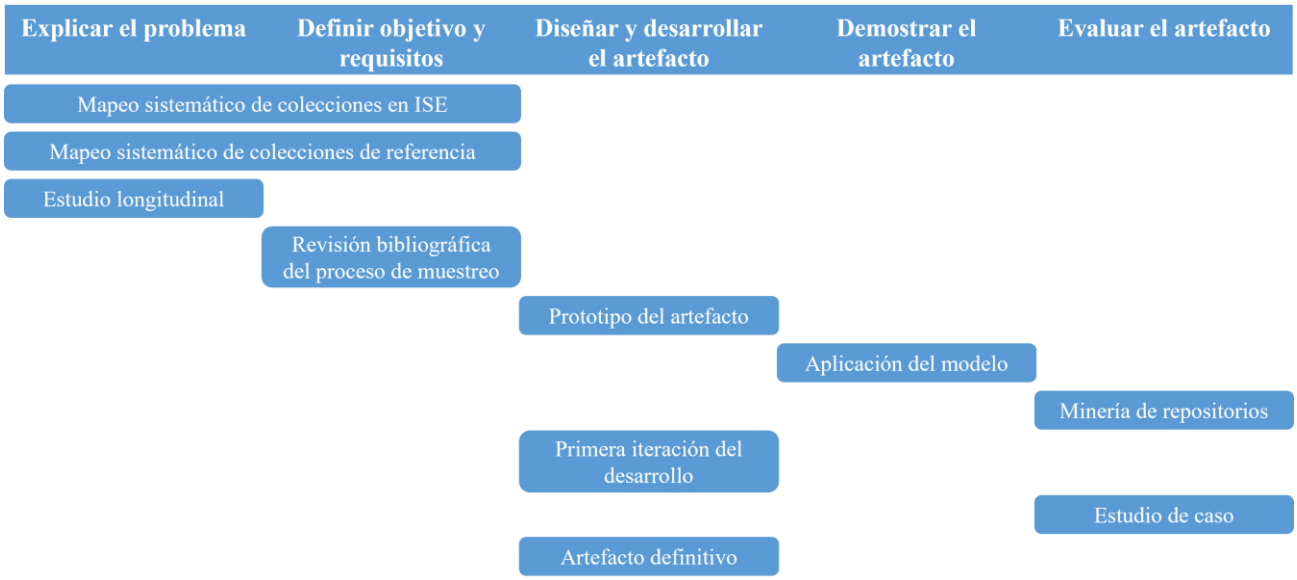


Fig. 1. Cronología de ejecución de las tareas de ciencia de diseño

En la mayoría de los estudios relevados se incluyeron criterios de selección específicos con el caso de estudio, demostrando la falta de estrategias de selección estandarizadas. Casi en su totalidad no consideraron utilizar la información más reciente para conducir los estudios, tanto es así que la mayoría de las colecciones nunca actualizaron las muestras (ver Fig. 2). Adicionalmente, se recopilaron los siguientes criterios para identificar proyectos aptos para la investigación en ISE: la disponibilidad del código fuente, la utilización de herramientas para dar soporte a procesos, el tamaño, la historia de desarrollo, la popularidad, la colaboración del equipo de desarrollo y la actividad reciente. Con respecto a los metadatos, se extrajeron métricas de código, pruebas unitarias, datos de repositorio, entre otras. De los cuales se emplearon métricas del código fuente y metadatos del repositorio para probar y validar SUM4SOFT.

En síntesis, al igual que otros estudios secundarios se evidenció la dificultad de acceso a los datos, reporte de herramientas y paquetes de replicación limitado, desestimación de la vigencia de las muestras y falta de estrategias para actualizar los datos.

Vigencia de las colecciones

La Fig. 2 refleja como la mayoría de las colecciones de referencia nunca han sido actualizadas y sus últimas versiones datan de 1983 a 2016. Esto podría presentar inconvenientes debido al proceso dinámico de evolución continua del software [15]. Además de la implementación de nuevas funcionalidades, corrección de errores, comprobación de funcionalidades, documentación, etc.; las tecnologías, entornos y bibliotecas utilizados para construir, probar y desplegar las soluciones informáticas también evolucionan con el tiempo, provocando a su vez la actualización del propio proyecto para mantenerse compatible.

La vigencia refiere a la medida en que una muestra representativa recopilada en el pasado continúa representando a la misma población en el presente, y cobra notoriedad en aquellos dominios donde los datos sufren cambios frecuentes [16]. Por ello, es necesario monitorear los datos recopilados en la colección de proyectos software para asegurar que siga representando a la población objetivo.

Para investigar la relevancia de la vigencia en colecciones de la ISE se analizó la evolución de una población de proyectos software durante un período de seis años (desde julio del 2017 hasta junio del 2023) a través de un estudio longitudinal [17]. El estudio abarcó siete métricas de repositorio: número de confirmaciones, colaboradores, PR cerradas, PR combinadas, incidentes resueltos, estrellas y bifurcaciones.

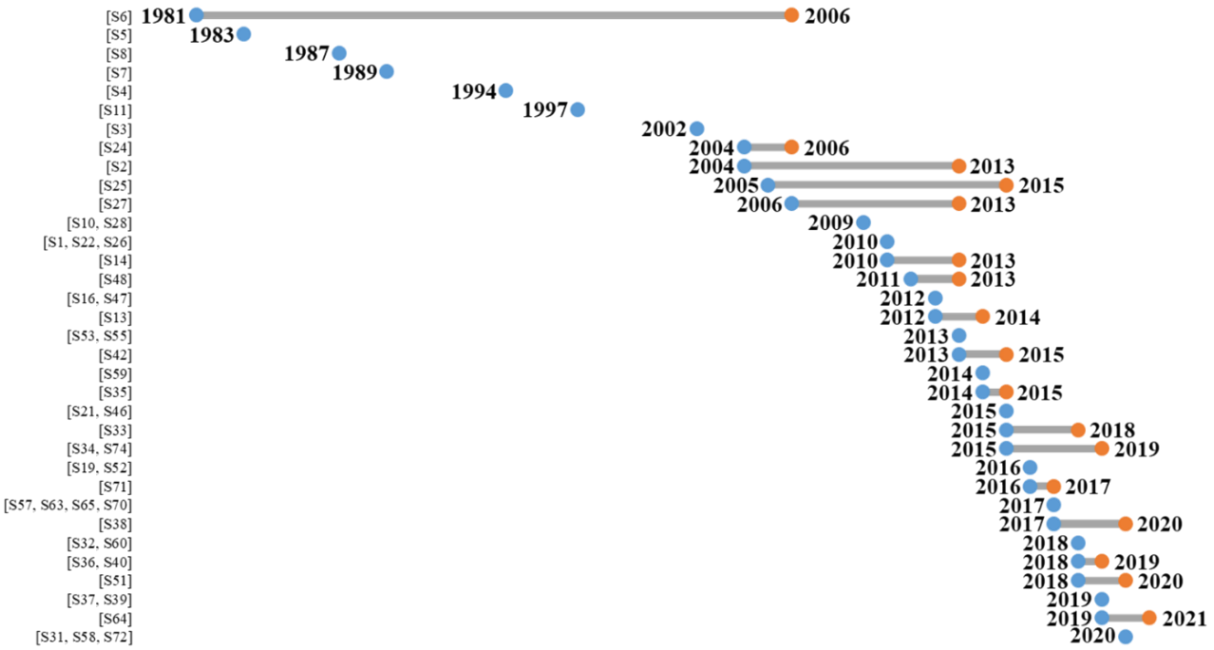


Fig. 2. Fecha de creación y última actualización de las colecciones de acceso público. Fuente [14]

El análisis estadístico realizado con la prueba de Kruskal-Wallis H [18] reveló diferencias significativas en todas las métricas en el período comprendido. En la Fig. 3 se graficaron las curvas de supervivencia de las métricas mostrando que, en la mayoría de ellas, la probabilidad de que una muestra con estos datos continúe siendo representativa por seis años o más es de 0%. Aunque la evolución no se manifiesta de manera uniforme en todas ellas, emplear conjuntos de datos creados hace varios años no parece ser una estrategia viable si el objetivo es producir resultados generalizables. Después de cinco años, todas las variables mostraron probabilidades de representatividad por debajo del 25%.

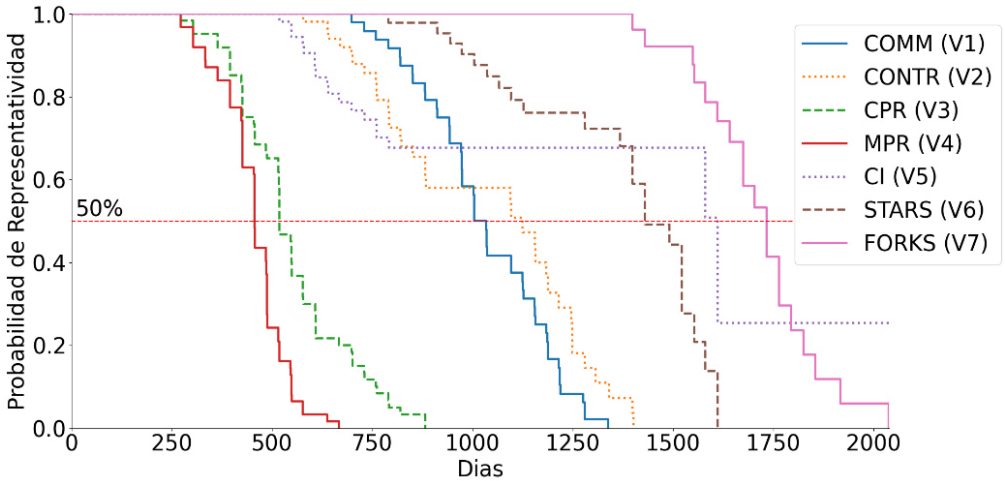


Fig. 3. Análisis de supervivencia de la representatividad dimensional. Fuente [17]

Desarrollo de SUM4SOFT

Según Marbán et al. [19] un modelo de proceso es un conjunto de pasos o actividades para llegar a un objetivo determinado. Estas actividades reciben las entradas necesarias para su ejecución y producen salidas. Por lo tanto, es necesario detallar cuáles son las actividades involucradas en los procesos modelados, lo que implica la especificación de sus entradas, salidas, orden de ejecución, y actores, entre otras cuestiones. Para la definición de las actividades se estudiaron las guías de muestreo en Ingeniería del Software relevadas por Baltes y Ralph [8].

SUM4SOFT (por sus siglas en inglés, ***S**ampling and **U**pdate **M**odel **for Soft**ware datasets*) es el modelo de proceso propuesto en esta tesis para abordar las problemáticas reportadas de los estudios en la ISE. El diseño del modelo abarca las cuatro perspectivas propuestas por Curtis et al. [20]: comportamiento, funcional, organizacional e informacional; y se diagramaron con el lenguaje de modelado SPEM 2.0. En la Fig. 4 se muestra la perspectiva de comportamiento del modelo completo donde se distinguen las dos actividades centrales: construir el instrumento de recolección de datos (A1) y ejecutar el instrumento de recolección de datos (A2).

La actividad A1 establece las tareas necesarias para generar el instrumento con el cual construir y actualizar las muestras. Las salidas de la actividad son un instrumento de recolección de datos acompañado de la documentación de diseño y las instrucciones de uso. La misma comprende las subactividades: especificar las fuentes (A1.1) y diseñar la estrategia de muestreo (A1.2); e implica tareas como buscar los recursos que proporcionarán los datos, caracterizar la población objetivo, diseñar la estrategia de muestreo, entre otras. El objetivo de la subactividad A1.1 es obtener una lista con las fuentes de los datos a utilizar y una guía simple de acceso a las mismas. Por su parte, el objetivo de la subactividad A1.2 es diseñar las reglas y estrategias para construir y conservar la colección curada de proyectos software partiendo de la guía de acceso a fuentes de la subactividad anterior.

La actividad A2, a partir del instrumento de recolección de datos creado en la A1, tiene como objetivo crear una colección de proyectos software y mantenerla vigente pese a los cambios producidos por el paso del tiempo.

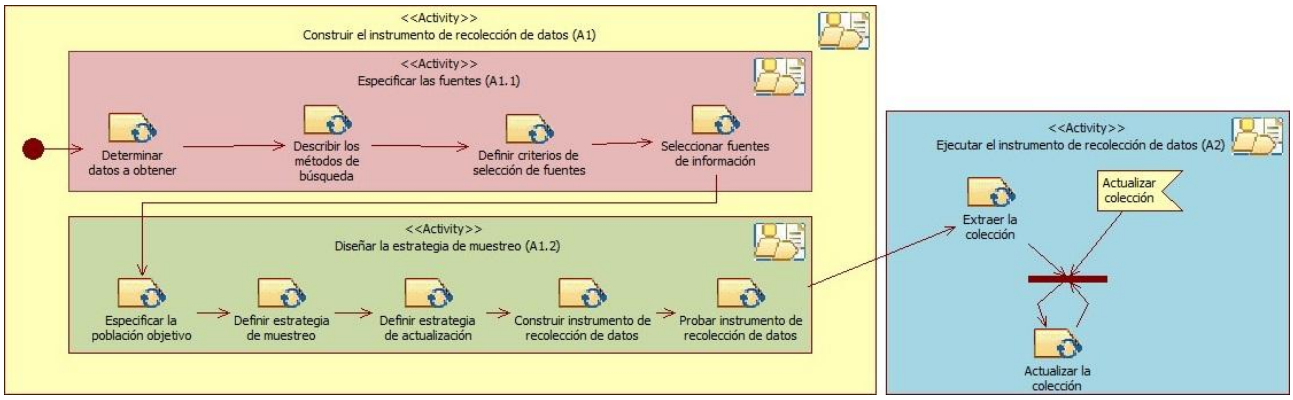


Fig. 4. Perspectiva de comportamiento de SUM4SOFT

Adicionalmente, se definieron tres roles de acuerdo con las competencias técnicas necesarias para desempeñar las actividades y tareas del modelo propuesto. Primero, el responsable de la colección: experto del dominio responsable de dirigir y supervisar todas las etapas del proceso, quien conoce en profundidad el tema de estudio, permitiéndole satisfacer los objetivos de la investigación. Segundo, el diseñador de los instrumentos de recolección de datos: encargado de las tareas de la subactividad A1.1, y además de construir y probar el instrumento de recolección de datos. Tercero, el analista de la muestra: cuenta con conocimientos en estadística aplicada necesarios para determinar las estrategias de muestreo y actualización, y verificar la composición de las colecciones.

Validación

Con el objetivo de demostrar la utilidad de SUM4SOFT en un contexto real de la ISE, se siguió el marco propuesto por el modelo para construir el instrumento de recolección de datos JavaQ.

En primer lugar, se especificaron los datos requeridos y dónde obtenerlos. A partir de los resultados de los mapeos sistemáticos, se identificaron los metadatos que cuantifican los criterios relevados y se obtuvieron dos posibles plataformas que brindaban acceso a estos datos: Github y Gitlab. Debido a

las limitaciones de Gitlab para el filtrado de proyectos por lenguaje de programación desde su servicio de consultas API REST, solo se recurrió a Github como fuente de los proyectos.

En segundo lugar, se definieron la población objetivo y las estrategias de muestreo y actualización. La población objetivo se conformó con aquellos proyectos que cumplieran con los umbrales de la Tabla 1, los cuales operacionalizan los criterios de selección. La lista de proyectos extraída sirvió de marco muestral para recolectar los proyectos aptos para la investigación.

Tabla 1. Umbrales para caracterizar la población objetivo

Umbral	**Criterio**
Lenguaje Java	Proyectos Java
Repositorios públicos	Metadatos disponibles
Repositorios GIT 50 incidentes o más	Uso de herramientas para soporte a procesos
Repositorio no copiado	Proyectos originales
10000 KB o más	Tamaño
3 o más colaboradores 50 PR o más	Colaboración
1 año desde creación o más 1000 confirmaciones o más	Historia
10 bifurcaciones o más 10 estrellas o más	Popularidad
1 confirmación/PR o más en el último mes 1 confirmación o más por mes en el último año Repositorio no archivado	Actividad reciente

En cuanto a la estrategia de muestreo, se empleó muestreo aleatorio estratificado para seleccionar los sujetos en el marco muestral al igual que estudios relacionados [21]. Esta técnica divide la población en subpoblaciones no solapadas denominados estratos y obtiene los elementos de cada estrato de manera aleatoria. La implementación del procedimiento de selección involucró la generación de los estratos a partir del tamaño del proyecto utilizando el algoritmo para agrupamiento k-means [22]. Antes de aplicar el algoritmo, se separaron los proyectos con valores extremos de tamaño y se aplicó k-means con cinco grupos. El número de grupos se fijó después inspeccionar el conjunto de datos completo y experimentar con diferentes valores. Una vez obtenidos los grupos, se procedió a seleccionar aleatoriamente los proyectos de cada grupo proporcionalmente de acuerdo con el tamaño del estrato y del marco muestral. A la muestra resultante, se le incorpora un subconjunto proporcional de los proyectos separados inicialmente por ser identificados como valores extremos. La Fig. 5 ilustra gráficamente la estrategia.

Con respecto a la estrategia de actualización, el proceso inicia generando un marco muestral nuevo a partir de los umbrales definidos en la Tabla 1. Si los elementos de la muestra original tienen actividad reciente se procede a actualizar los proyectos en la muestra por las nuevas versiones disponibles. En cambio, los proyectos sin actualizaciones se descartan de la muestra actualizada. Los proyectos descartados son reemplazados a partir de un proceso similar al muestreo estratificado aleatorio. Primero, se generan los estratos a partir marco muestral nuevo de la misma manera que la estrategia de muestreo, es decir, se separan lo valores extremos y se aplica k-means con cinco grupos para crear los estratos. Segundo, se registra la cantidad de proyectos de cada estrato que existen en la muestra actualizada (sin los proyectos descartados) y se obtiene la diferencia con la cantidad proporcional de elementos correspondientes al estrato equivalente en el marco muestral. Tercero, se extraen los proyectos en el marco muestral de acuerdo con las diferencias registradas por grupo en el paso previo.

Esta herramienta se empleó en dos estudios empíricos: un estudio de minería de repositorios y un estudio de caso. El objetivo del estudio de minería de repositorios [23] fue evaluar la efectividad de JavaQ para construir y actualizar colecciones de proyectos. Para ello, se generaron los umbrales de

calidad de tres métricas de código: complejidad ciclomática (McCC), cantidad de líneas de código (LOC) y peso de métodos por clase (WMC). Los umbrales de métricas delimitan puntos de referencia aceptables y anormales en los componentes de software, proporcionando un indicador para evaluar el estado del sistema [24]. Para generar los umbrales de las métricas se implementó el enfoque estadístico basado en datos planteado por Alves et al. [25]. Esta técnica calcula la función de distribución acumulada de la métrica estudiada y establece tres valores de referencia: límite inferior (70%), medio (80%) y límite superior (90%), delimitando cuatro niveles de riesgo de un componente de sistema: riesgo bajo (<= 70%), riesgo moderado (70% - 80%), riesgo alto (80% - 90%) y riesgo muy alto (> 90%).

Se utilizaron tres colecciones para la evaluación: una muestra desactualizada (Qualitas Corpus), una muestra con datos recientes (Colección Vigente) y una versión nueva de la muestra desactualizada (Qualitas Actualizado). La muestra desactualizada corresponde a la versión publicada en 2013 de la colección Qualitas Corpus [26], la colección vigente se obtuvo de aplicar la estrategia de muestreo estratificado, y la colección Qualitas actualizado es una versión actualizada del Qualitas Corpus con la estrategia de actualización.

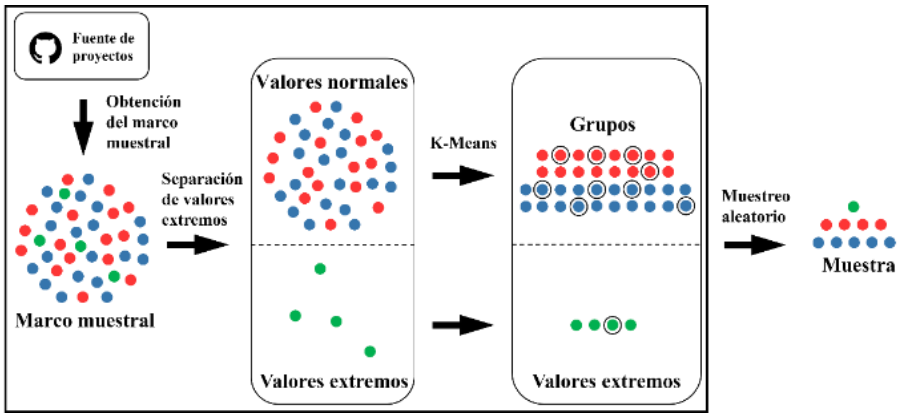


Fig. 5. Esquema de la estrategia de muestreo

Para el análisis estadístico se utilizó la prueba de hipótesis Kolmogorov-Smirnov de dos muestras independientes [27]. Esta prueba evalúa si dos muestras provienen de la misma población, y si ese fuese el caso, se espera que sus distribuciones de frecuencias acumuladas sean idénticas o muy cercanas entre sí. Dicho esto, la prueba reveló diferencias significativas entre la colección vigente y el Qualitas Corpus en la distribución de probabilidades de las métricas LOC, McCC y WMC generando dudas acerca de la validez del Qualitas Corpus para conducir investigaciones actuales centradas en estas propiedades. En cambio, la comparación de los umbrales obtenidos de las colecciones recolectadas con las estrategias de muestreo y actualización implementadas en JavaQ presentaron diferencias marginales. En la Fig. 6 se trazaron los umbrales de las tres muestras analizadas en el estudio, donde las curvas que corresponden a las colecciones obtenidas de JavaQ (azul y naranja) prácticamente se superponen de principio a fin, y la verde (Qualitas Corpus) tiene una distancia considerable de las demás.

Para validar los resultados obtenidos en el estudio de minería de repositorios y recibir mejoras para el modelo de proceso se realizó un estudio de caso siguiendo los lineamientos establecidos por Runeson et al. [28]. El estudio se desarrolló en la Universidad de Hamburgo con el grupo de investigación *Software Engineering and Construction Methods* [29] en el marco de una estancia de investigación doctoral[1].

[1] Las Becas ALEARG son resultado del convenio firmado entre el Ministerio de Capital Humano de la República Argentina y el Servicio Alemán de Intercambio Académico para realizar estancias cortas de investigación en universidades alemanas en el marco de estudios doctorales o posdoctorales en universidades argentinas.

En la primera etapa, se condujo una replicación del estudio de minería de repositorios enmarcado en el estudio de caso aplicando las estrategias de muestreo implementadas en JavaQ. Al igual que en el estudio original, los investigadores externos extrajeron dos colecciones con proyectos actuales de Github: una colección vigente y una versión actualizada del Qualitas Corpus. A partir de ellos se derivaron los umbrales de las métricas LOC, McCC y WMC, y se contrastaron con los obtenidos en el estudio original. El análisis estadístico de los umbrales de referencia obtenidos en la replicación y el estudio original mostraron resultados consistentes para las métricas LOC, McCC, y WMC. Se observaron diferencias mínimas entre las colecciones junto con p-valores mayores a 0.85 en la prueba Kolmogorov-Smirnov, sugiriendo que las distribuciones de los datos en ambos estudios provienen de la misma población. Este hallazgo confirmó la validez del instrumento de recolección de datos JavaQ, aportando robustez a los resultados originales y habilitando la generalización a contextos similares.

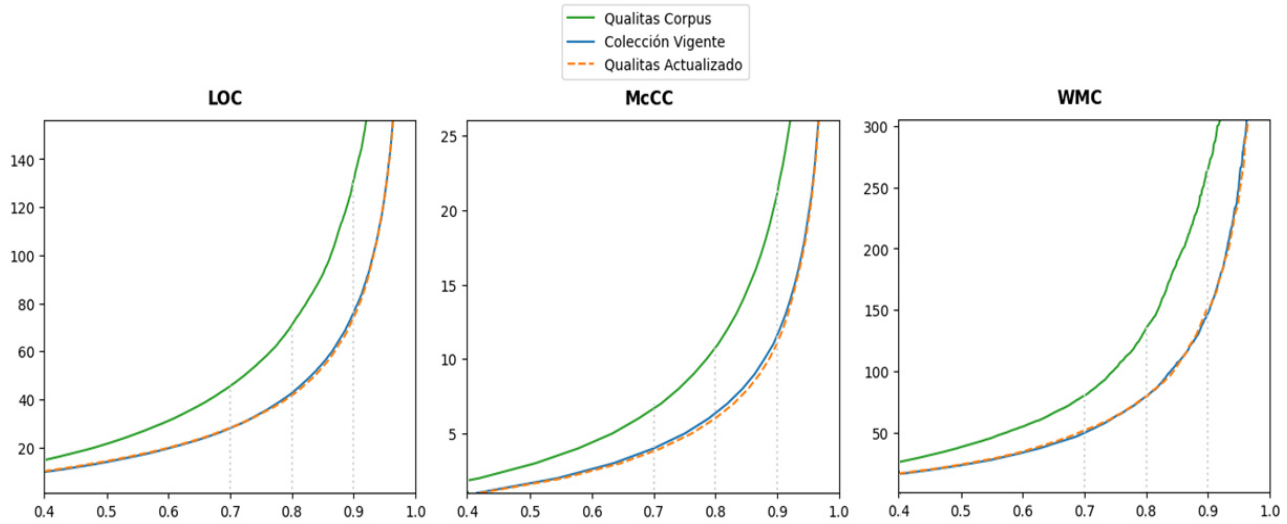


Fig. 6. Funciones de distribución acumulada de cada colección

El estudio de caso también tuvo como objetivo capturar sugerencias del grupo de investigación externo para refinar las tareas y roles del modelo de proceso propuesto. Las respuestas se analizaron cuantitativa y cualitativamente y se sintetizaron en la Tabla 2. Esto permitió operacionalizar el perfeccionamiento de las tareas y roles de SUM4SOFT, los cuales fueron incluidos en la versión definitiva del artefacto para facilitar su uso en grupos externos.

Tabla 2. Modificaciones de tareas incluidas en el modelo

Act.	Tarea	Motivación
A1.1	Describir los métodos de búsqueda	Agregar métodos de búsqueda
A1.2	Especificar la población objetivo	Instrumentar criterios de selección
	Definir estrategia de muestreo	Informar sobre el tamaño de la muestra
	Definir estrategia de actualización	Estandarizar estrategia de actualización
	Probar instrumento de recolección	Verificar el instrumento
A2	Actualizar la colección	Comprobar desviaciones

Contribuciones

La contribución principal de esta investigación doctoral fue el desarrollo de herramientas para el soporte de estudios en la Ingeniería del Software Empírica. En primer lugar, el modelo de proceso para la construcción y actualización de muestras de proyectos software SUM4SOFT. En segundo lugar, las salidas del modelo obtenidos en el contexto del estudio de minería de repositorios: una colección de proyectos software y el instrumento de recolección de datos JavaQ con las herramientas para mantener la colección facilitando la replicación de los resultados. Tanto los datos como JavaQ están disponibles en la plataforma para datos abiertos Zenodo[2].

También se profundizó acerca del estado actual de las colecciones de proyectos software utilizadas en la ISE a través de tres estudios empíricos: dos mapeos sistemáticos y un estudio longitudinal. Las revisiones de la literatura confirmaron algunas problemáticas mencionadas por otros autores como: la falta de estrategias de selección de proyectos estandarizadas, la omisión de la vigencia como criterio de selección, la ausencia del muestreo probabilístico y las restricciones de acceso, tanto a las colecciones como a las herramientas empleadas en la experimentación con ellas. A su vez se identificaron las características deseables en los proyectos software para construir una colección curada. Por su parte, el estudio longitudinal midió el impacto del tiempo en una población de proyectos y demostró la importancia de revisar la vigencia de las colecciones en contextos donde la población evoluciona de manera constante.

Dentro de los artículos publicados de la investigación se encuentran: el simposio doctoral [11], los mapeos sistemáticos de la literatura [13, 14], y el estudio longitudinal [17]. A continuación, se enumeran los estudios pendientes de publicación:

- Estudio de minería de repositorios. El estudio demostró la utilidad de JavaQ, instrumento obtenido de SUM4SOFT para generar colecciones representativas. El manuscrito se envió a la revista *Information and Software Technology* y se encuentra en proceso de revisión de pares.
- Estudio de caso. Este artículo presenta el modelo de proceso SUM4SOFT y las evaluaciones conducidas durante la estancia doctoral en la Universidad de Hamburgo. El manuscrito se encuentra en redacción y será enviado a la revista *Software and Systems Modeling*.

Trabajos futuros

Para continuar con trabajos que contribuyan a la línea actual se propone validar SUM4SOFT en otros contextos. Si bien el modelo de proceso se validó inicialmente en un estudio de minería y posteriormente en un estudio de caso utilizando el instrumento de recolección de datos JavaQ para generar las colecciones, probarlo en un contexto de aplicación diferente al de métricas del código fuente le proporcionaría mayor robustez. El objetivo consistirá en monitorear la utilización de SUM4SOFT en un nuevo estudio empírico diseñado por investigadores externos para verificar su aplicabilidad en una variedad más amplia de escenarios.

Por otra parte, dado que la pérdida de la vigencia de las colecciones de software es un problema que afecta no solo a las métricas de código fuente y de repositorio, sino a los proyectos en su conjunto, resulta lógico extender el estudio de la vigencia a otros tópicos de la Ingeniería del Software. Por ejemplo, las pruebas del software, la seguridad, refactorización del código, entre otras. De esta manera, se puede concientizar a los investigadores y profesionales de los riesgos de trabajar con muestras no vigentes.

[2] https://doi.org/10.5281/zenodo.11059968.

Referencias

[1] B. Kitchenham, T. Dyba, and M. Jorgensen, "Evidence-based software engineering," in *Proceedings. 26th International Conference on Software Engineering*, IEEE Comput. Soc, 2004, pp. 273–281. doi: 10.1109/ICSE.2004.1317449.

[2] C. Wohlin and A. Aurum, "Towards a decision-making structure for selecting a research design in empirical software engineering," *Empir Softw Eng*, vol. 20, no. 6, pp. 1427–1455, Dec. 2015, doi: 10.1007/s10664-014-9319-7.

[3] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén, *Experimentation in Software Engineering*, vol. 9783642290. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. doi: 10.1007/978-3-642-29044-2.

[4] E. Barr, J. Bell, M. Hilton, S. Mechtaev, and C. Timperley, "Continuously Accelerating Research," in *2023 IEEE/ACM 45th International Conference on Software Engineering: New Ideas and Emerging Results (ICSE-NIER)*, IEEE, May 2023, pp. 123–128. doi: 10.1109/ICSE-NIER58687.2023.00028.

[5] M. M. Lehman, "Laws of software evolution revisited," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Verlag, 1996, pp. 108–124. doi: 10.1007/BFb0017737.

[6] N. Munaiah, S. Kroh, C. Cabrey, and M. Nagappan, "Curating GitHub for engineered software projects," *Empir Softw Eng*, vol. 22, no. 6, pp. 3219–3253, Dec. 2017, doi: 10.1007/s10664-017-9512-6.

[7] E. Kalliamvakou, G. Gousios, K. Blincoe, L. Singer, D. M. German, and D. Damian, "The promises and perils of mining GitHub," in *Proceedings of the 11th Working Conference on Mining Software Repositories*, New York, NY, USA: ACM, May 2014, pp. 92–101. doi: 10.1145/2597073.2597074.

[8] S. Baltes and P. Ralph, "Sampling in software engineering research: a critical review and guidelines," *Empir Softw Eng*, vol. 27, no. 4, p. 94, Jul. 2022, doi: 10.1007/s10664-021-10072-8.

[9] V. Cosentino, J. Luis, and J. Cabot, "Findings from GitHub," in *Proceedings of the 13th International Conference on Mining Software Repositories*, New York, NY, USA: ACM, May 2016, pp. 137–141. doi: 10.1145/2901739.2901776.

[10] P. Johannesson and E. Perjons, *An Introduction to Design Science*, vol. 9783319106. Cham: Springer International Publishing, 2014. doi: 10.1007/978-3-319-10632-8.

[11] J. A. Carruthers, "Open-Source Software Projects Curating Model for Empirical Software Engineering Studies," in *Anais do XXV Congresso Ibero-Americano em Engenharia de Software (CIbSE 2022)*, Sociedade Brasileira de Computação, Jun. 2022, pp. 416–423. doi: 10.5753/cibse.2022.20992.

[12] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, "Systematic Mapping Studies in Software Engineering," in *Proceedings of the 12th international conference on Evaluation and Assessment in Software Engineering*, 2008, pp. 68–77. doi: 10.14236/ewic/EASE2008.8.

[13] J. A. Carruthers, J. A. Diaz-Pace, and E. A. Irrazabal, "A Systematic Mapping Study of Empirical Studies Performed with Collections of Software Projects," *Computación y Sistemas*, vol. 26, no. 4, Dec. 2022, doi: 10.13053/cys-26-4-3981.

[14] J. A. Carruthers, J. A. Diaz-Pace, and E. A. Irrazabal, "How are software datasets constructed in Empirical Software Engineering studies? A systematic mapping study," in *2022 48th Euromicro*

Conference on Software Engineering and Advanced Applications (SEAA), IEEE, Aug. 2022, pp. 442–450. doi: 10.1109/SEAA56994.2022.00075.

[15] M. M. Lehman, "Programs, life cycles, and laws of software evolution," *Proceedings of the IEEE*, vol. 68, no. 9, pp. 1060–1076, 1980, doi: 10.1109/PROC.1980.11805.

[16] K. Munger, "The Limited Value of Non-Replicable Field Experiments in Contexts With Low Temporal Validity," *Soc Media Soc*, vol. 5, no. 3, p. 205630511985929, Apr. 2019, doi: 10.1177/2056305119859294.

[17] J. A. Carruthers, J. A. Diaz-Pace, and E. Irrazábal, "A longitudinal study on the temporal validity of software samples," *Inf Softw Technol*, vol. 168, p. 107404, Apr. 2024, doi: 10.1016/j.infsof.2024.107404.

[18] W. H. Kruskal and W. A. Wallis, "Use of Ranks in One-Criterion Variance Analysis," *J Am Stat Assoc*, vol. 47, no. 260, pp. 583–621, Dec. 1952, doi: 10.1080/01621459.1952.10483441.

[19] Ó. Marbán, G. Mariscal, E. Menasalvas, and J. Segovia, "An engineering approach to data mining projects," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 4881 LNCS, pp. 578–588, 2007, doi: 10.1007/978-3-540-77226-2_59/COVER.

[20] B. Curtis, M. I. Kellner, and J. Over, "Process modeling," *Commun ACM*, vol. 35, no. 9, pp. 75–90, Sep. 1992, doi: 10.1145/130994.130998.

[21] J. Gorostidi, A. Ait, J. Cabot, and J. L. Canovas Izquierdo, "On the Creation of Representative Samples of Software Repositories," in *Proceedings of the 18th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*, New York, NY, USA: ACM, Oct. 2024, pp. 434–439. doi: 10.1145/3674805.3690747.

[22] Y. Bugayenko *et al.*, "Qualitative Clustering of Software Repositories Based on Software Metrics," *IEEE Access*, vol. 11, pp. 14716–14727, 2023, doi: 10.1109/ACCESS.2023.3244495.

[23] M. Vidoni, "A systematic process for Mining Software Repositories: Results from a systematic literature review," *Inf Softw Technol*, vol. 144, p. 106791, Apr. 2022, doi: 10.1016/j.infsof.2021.106791.

[24] M. Lanza and R. Marinescu, *Object-Oriented Metrics in Practice*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006. doi: 10.1007/3-540-39538-5.

[25] T. L. Alves, C. Ypma, and J. Visser, "Deriving metric thresholds from benchmark data," in *2010 IEEE International Conference on Software Maintenance*, IEEE, Sep. 2010, pp. 1–10. doi: 10.1109/ICSM.2010.5609747.

[26] E. Tempero *et al.*, "The Qualitas Corpus: A Curated Collection of Java Code for Empirical Studies," in *2010 Asia Pacific Software Engineering Conference*, IEEE, Nov. 2010, pp. 336–345. doi: 10.1109/APSEC.2010.46.

[27] D. J. Sheskin, *Parametric and Nonparametric Statistical Procedures*, 2nd ed. 2000.

[28] P. Runeson, M. Höst, A. Rainer, and B. Regnell, *Case Study Research in Software Engineering*. Wiley, 2012. doi: 10.1002/9781118181034.

[29] S. S. Hemming, "Software Engineering and Construction Methods Research Group," University of Hamburg, 2023. https://www.inf.uni-hamburg.de/en/inst/ab/swk/home.html.

Posibilidades Pedagógicas de los Entornos Virtuales 3D en el Acompañamiento del Aprendizaje de Personas con Discapacidad Auditiva

Adriana Silvia Fachal[1][0000-0002-6379-2856], María José Abásolo[1][0000-0003-4441-3264], Cecilia Sanz[1][0000-0002-9471-0008]

[1] Instituto de Investigación en Informática LIDI - CIC, Facultad de Informática – Universidad Nacional de La Plata, La Plata, Argentina
afachal@hotmail.com, mjabasolo@ lidi.info.unlp.edu.ar , csanz@lidi.info.unlp.edu.ar

Fecha de Exposición: 14/10/2024

1. Motivación

Este trabajo se vincula con la experiencia personal de la tesista como docente donde identificó la falta de apoyo para estudiantes con discapacidad auditiva.
A lo largo de 18 años de enseñanza, la tesista se ha formado como Intérprete en Lengua de Señas Argentina (LSA) y ha utilizado Tecnologías de la Información y Comunicación (TIC) para fortalecer la educación bilingüe de estos estudiantes.
La motivación central de esta tesis radica en mejorar la inclusión educativa de las personas con discapacidad auditiva, quienes enfrentan barreras comunicativas en un contexto mayoritariamente oyente.
La propuesta se enfoca en el uso de Entornos Virtuales 3D (EV3D), que facilitan el aprendizaje, la socialización y la autonomía de estos estudiantes mediante la creación de espacios interactivos y colaborativos. Esta investigación busca explorar las potencialidades educativas y de integración social de los EV3D, con el objetivo de fomentar la participación activa y mejorar los resultados de aprendizaje en un entorno accesible y motivador.

Palabras clave: educación inclusiva; discapacidad auditiva; e-accesibilidad; TIC (Tecnologías de la Información y la Comunicación) ; mundos virtuales ; metaverso ; entornos virtuales 3D ; OpenSimulator

2. Introducción

Las personas con discapacidad auditiva enfrentan barreras comunicativas que dificultan su acceso a la información y la comunicación, especialmente en los contextos educativos y sociales, lo que impacta su desarrollo personal, social y laboral. La educación, como factor clave en el desarrollo de habilidades para una vida profesional y social, es crucial para su integración a la sociedad [1]. Un modelo educativo bilingüe-cultural, que promueva el dominio tanto de la Lengua de Señas (LS) como del idioma oficial, es fundamental para la inclusión de este colectivo [2].

El uso de Tecnologías de la Información y la Comunicación (TIC) ha demostrado ser una herramienta eficaz para mejorar los procesos de enseñanza y aprendizaje de personas con discapacidad auditiva. En particular, los EV3D ofrecen un espacio interactivo y multisensorial que facilita la participación de los estudiantes en actividades educativas a través de avatares, favoreciendo tanto la comunicación verbal como no verbal [3]. Estas características inmersivas potencian la motivación de los estudiantes y promueven un aprendizaje activo y colaborativo [4].

Este estudio se centra en cómo los EV3D pueden mejorar la experiencia educativa de las personas con discapacidad auditiva, promoviendo la participación, la expresión de su identidad y el aprendizaje por descubrimiento. A través de la

implementación de estos entornos virtuales, se busca ofrecer una solución innovadora que impulse el acceso a la educación inclusiva y la integración social de este grupo [5].

3. Metodología de la Investigación

La metodología adoptada para la tesis es de carácter mixto, combinando enfoques cualitativos y cuantitativos. Se desarrolló en tres etapas principales:

1. Investigación teórica: Se llevó a cabo una investigación exhaustiva sobre los conceptos clave relacionados con la comunicación y educación en personas con discapacidad auditiva, así como el uso de herramientas digitales (TIC) en este contexto. Se abordó el concepto de EV3D y sus posibilidades educativas. Esta etapa proporcionó el marco teórico fundamental para la tesis.

2. Revisión de antecedentes: En esta etapa se realizó una revisión sistemática de la literatura para analizar las experiencias previas de uso de EV3D en la educación y comunicación de personas con discapacidad auditiva. Este análisis permitió identificar buenas prácticas y áreas de oportunidad en el uso de estos entornos para la inclusión educativa.

3. Desarrollo de los aportes de la tesis: En la fase final, se diseñaron guías de diseño para crear un EV3D inclusivo para personas con discapacidad auditiva, y una metodología en etapas para guiar el desarrollo de una experiencia educativa dentro de un EV3D. Se utilizó la plataforma OpenSim[1] para implementar los escenarios y herramientas, como un tablero de emojis, empleando metodologías ágiles con reuniones periódicas de seguimiento, prototipos evolutivos y testeos continuos. Además, se llevó a cabo un estudio de caso con grupos presenciales y a distancia (en colaboración con el IIDE y asociaciones como ALPI y ASAM), para evaluar aspectos relacionados con la comunicación (uso de chat y el tablero de emojis), identidad (a través del avatar) y motivación de los participantes en relación con las actividades educativas realizadas.

4. Marco Teórico y Estado del Arte

Esta sección está compuesta sobre 2 ejes, la Discapacidad Auditiva y los EV3D

4.1. Discapacidad Auditiva

La discapacidad auditiva es una alteración anatómica y/o fisiológica del sistema auditivo que impacta el acceso al lenguaje oral [6].

Puede manifestarse como sordera profunda o hipoacusia [7], clasificándose según su intensidad, localización y momento de aparición. De acuerdo con el Bureau Internacional de Audiología, la pérdida auditiva varía desde audición normal (0-20 dB) hasta la cofosis o anacusia, que representa una pérdida total de la audición (mayor a 90 dB).

Las personas con discapacidad auditiva emplean la lengua de señas (LS), una lengua gesto-visual que no es universal ya que varía según la comunidad lingüística, como por ejemplo la Lengua de Señas Argentina (LSA).

En este contexto, una persona con discapacidad auditiva es considerada bilingüe si tiene habilidad tanto en Lengua de Señas como en la lengua oficial del país (en este caso, el español).

Este panorama ha generado un debate en el ámbito educativo sobre el modelo de enseñanza más adecuado para las personas con discapacidad auditiva. Se discuten dos enfoques principales: la integración en centros ordinarios (donde estudiantes sordos asisten a escuelas comunes junto a oyentes) y los centros educativos especiales dedicados exclusivamente a personas con discapacidad auditiva. Además, se plantea la discusión sobre los modelos educativos monolingües frente a los bilingües. Además al trabajar con tecnologías es importante considerar metodologías como la M-free que apoya la planificación de una actividad educativa dirigida a alumnos con necesidades específicas.

[1] http://opensimulator.org/

4.2. Entornos Virtuales 3D (EV3D)

En el marco de esta tesis, el concepto de EV3D ha sido abordado desde la perspectiva de las definiciones propuestas por diversos autores [9][10][11] [12][13]. A partir de este análisis, se adoptó la siguiente definición:

- Los EV3D son representaciones digitales 3D accesibles a través de internet que permiten simular un mundo o entorno artificial inspirado o no en la realidad y donde los usuarios pueden interactuar a través de avatares y comunicarse con otros a través de mensajes instantáneos de texto y voz.

Esta definición surge del análisis de múltiples perspectivas teóricas que destacan distintos aspectos de los EV3D, tales como su capacidad de representación gráfica, su uso en la socialización y la educación, y su potencial para fomentar experiencias colaborativas y accesibles [8] [14].

Se destacan por sus características de ser inmersivos (sensación de estar dentro del EV3D), interactivos (comunicar e interactuar en tiempo real), personalizables (permite modificación de los escenarios) y persistentes (se sigue desarrollando aún off-line)

Entre las plataformas utilizadas para la simulación de EV3D, se destacan Second Life (SL) y OpenSimulator (OpenSim). SL, lanzada en 2003, es una comunidad virtual de código cerrado donde los usuarios interactúan mediante avatares y participan en una economía virtual controlada por Linden Labs. En contraste, OpenSim, disponible desde 2007, es una alternativa de código abierto que permite la creación y gestión de entornos virtuales personalizados, ofreciendo mayor control y flexibilidad, lo que la hace particularmente adecuada para el ámbito educativo [14][15].

Si bien los EV3D no fueron concebidos originalmente como plataformas educativas, su uso se ha expandido en este contexto, facilitando experiencias de aprendizaje inmersivas y colaborativas [17]. Su capacidad de interacción en tiempo real [16] y sensación de presencialidad aporta un valor pedagógico significativo en la educación a distancia [18]. En esta investigación, se ha optado por OpenSim debido a sus ventajas educativas, su flexibilidad y la posibilidad de crear entornos de aprendizaje colaborativo que favorecen la interacción y la presencia virtual [19].

4.3. EV3D y Discapacidad

Los EV3D presentan un gran potencial en la educación inclusiva, facilitando la integración de personas con diversas discapacidades mediante herramientas accesibles y adaptativas [20] [21]. En el caso de la discapacidad auditiva, estos entornos permiten la combinación de elementos visuales, textuales y tridimensionales para complementar el discurso oral [12] [17].

Un aspecto relevante es la personalización de avatares, que pueden reflejar o anonimizar la discapacidad de los usuarios. Mientras algunos defienden la representación explícita de la discapacidad como parte de su identidad, otros optan por un diseño neutro para favorecer la interacción social [8] [17]. Esta dualidad plantea el desafío de garantizar la inclusión respetando la autonomía del usuario [6].

5. Aportes de la Tesis

En esta sección se sintetizan los aportes centrales que surgen como parte de la propuesta de la tesis.

El aporte principal de esta tesis se organiza en tres ejes fundamentales que contribuyen a la creación de experiencias educativas inclusivas en EV3D para personas con discapacidad auditiva (Figura 01).

- Guías de diseño: Se proponen una serie de guías orientadas específicamente a personas con discapacidad auditiva, con el objetivo de generar experiencias de aprendizaje más accesibles y efectivas en entornos virtuales 3D.

- Metodología para el desarrollo educativo: Se define una metodología que consta de cinco etapas, y se basa en las guías de diseño aportadas, para guiar a docentes e instituciones en la creación de experiencias educativas dentro de entornos EV3D. Esta metodología integra aspectos pedagógicos y didácticos clave para promover la autonomía, la motivación, el trabajo colaborativo y la comunicación de los estudiantes.

- Decisiones de diseño para escenarios educativos: Se presentan las decisiones de diseño necesarias para la construcción de escenarios educativos adaptados a las necesidades de los usuarios. Estas decisiones están influenciadas tanto por las guías de diseño como por la metodología propuesta, y tienen como objetivo asegurar que los entornos virtuales favorezcan la integración social y el aprendizaje de las personas con discapacidad auditiva.

En resumen, esta tríada ha permitido crear un EV3D con escenarios desarrollados ad-hoc que se orientan a ayudar a las personas con discapacidad auditiva a mejorar su integración social, su comunicación y el aprendizaje y enseñanza.

Se diseña así una experiencia educativa para EV3D que se integró al estudio de caso de la tesis.



Figura 01. Los 3 ejes principales del Aporte de la Tesis (Elaboración Propia)

5.1. Guías, Metodología, Decisiones de Diseño para personas con discapacidad auditiva

Las guías desarrolladas surgen de la revisión del estado del arte y del marco teórico de la tesis y orientaron la creación de los escenarios virtuales 3D ad-hoc abordados para el estudio de caso de esta tesis. Las guías establecen características fundamentales para trabajar con EV3D, especialmente orientadas a personas con discapacidad auditiva. A continuación, se detallan estas guías de diseño:

- GD1. Crear un entorno virtual interactivo 3D con avatares sobre PC para permitir mayor participación y acceso, y evitar fatiga y mareos que pueden darse en experiencias con gafas.

- GD2. Aprovechar las representaciones visuales del entorno 3D, la manipulación, la creación, y la observación de objetos que sean relevantes para la experiencia y el aprendizaje.

- GD3. Proporcionar guías y apoyos para generar un sentimiento de autonomía al usuario para que pueda desarrollar acciones por sí mismo

- GD4. Ofrecer contenidos en múltiples formatos multimediales

- GD5. Realizar una planificación pedagógica para llevar adelante actividades dentro del EV3D, se toma en consideración el modelo didáctico M-Free

- GD6. Problematizar la experiencia educativa en el EV3D para que los estudiantes sientan que el tema a abordar es relevante para ser resuelto o trabajado

- GD7. Proponer técnicas de gamificación fomentando el disfrute y la motivación de los estudiantes con desafíos dinámicos

- GD8. Fomentar el uso de diferentes recursos lingüísticos y de comunicación en función del estilo cognitivo del usuario (por ej. uso del chat textual y elementos visuales, como videos subtitulados)

- GD9. Proponer actividades que fomenten la comunicación y la socialización

- GD10. Utilizar lenguaje adecuado al público destinatario (en este caso LSA)

- GD11. Posibilitar la creación de una identidad en el EV3D a partir de la personalización del avatar, de manera que cada usuario sea quién decida cómo representarse

La metodología propuesta en esta tesis está diseñada para guiar la implementación de experiencias educativas en EV3D para personas con discapacidad auditiva. Este enfoque se estructura en cinco etapas clave, que constituyen el marco del estudio de caso (Figura 02).

- Etapa 1: Convocatoria y selección de participantes: En esta etapa inicial, se realiza la convocatoria y selección de los participantes con discapacidad auditiva.

- Etapa 2: Instalación del visor EV3D: Los participantes descargan e instalan el visor necesario para acceder al entorno EV3D.
- Etapas 3 a 5. Son tres etapas en las que se llevan a cabo sesiones específicas con actividades educativas dentro del EV3D. Estas equilibran el trabajo individual y grupal, e incluyen:
- Sesión 1 de adaptación e identidad: En esta etapa, se propone que el docente diseñe una actividad introductoria en la que los participantes se adapten al entorno virtual y creen su avatar, estableciendo su identidad dentro del EV3D.
- Sesión 2 de Exploración de contenidos y socialización: se sugiere realizar una actividad de exploración de contenidos, para que los participantes interactúen con el entorno, trabajando de manera individual y grupal, y participen en un debate en función de los temas educativos seleccionados.
- Sesión 3 con gamificación para el aprendizaje: Finalmente, se sugiere como parte de la metodología, implementar una actividad gamificada centrada en el aprendizaje de un tema específico, incentivando la participación y la interacción en el entorno virtual.

Esta metodología involucra a dos actores clave: el docente, quien utiliza la metodología para planificar e implementar las actividades, y los participantes, quienes intervienen a partir de la etapa 2, luego de recibir la convocatoria y aceptarla. Al mismo tiempo, sugiere un equilibrio entre actividades individuales y actividades grupales, permitiendo una interacción dinámica y colaborativa dentro del EV3D.

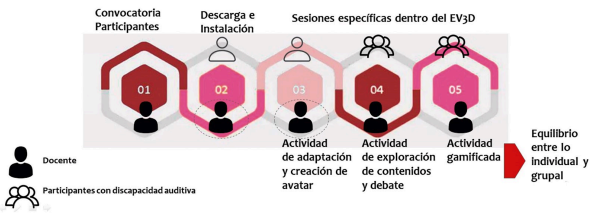


Figura 02. Metodología en 5 etapas para planificar e implementar (Elaboración Propia)

Además, se sugiere el uso de las guías de diseño propuestas en la tesis para tomar decisiones clave en cada una de las etapas, asegurando que se adapten a las necesidades educativas y de accesibilidad de los participantes.

La metodología es flexible y replicable, lo que permite su implementación en otros contextos educativos que deseen incorporar entornos EV3D para personas con discapacidad auditiva.

En esta tesis se creó utilizando OpenSim un EV3D y una experiencia educativa siguiendo la metodología y las guías de diseño propuestas. Tanto el diseño del EV3D como las etapas metodológicas, incluyendo las actividades y herramientas utilizadas, son propuestas originales de esta tesis y fueron validadas a través de su aplicación en el estudio de caso con diversos grupos de participantes.

Las principales decisiones para la creación del EV3D siguiendo las directrices de diseño mencionadas fueron: 1) Desarrollo de escenarios visuales e interactivos en OpenSim, 2) Utilización y personalización de avatares, 3) Cartelería y videotutoriales en LSA subtitulados, 4) Andamiajes previos, 5) Desarrollo de actividades educativas inmersivas y situadas con feedback inmediato, 6) Chat enriquecido con un tablero de emojis creado como parte de la tesis.

5.2. Diseño del EV3D. Modelado de los Escenarios Ad-hoc para el EV3D.

En el desarrollo de los escenarios para el EV3D, se llevó a cabo un análisis exhaustivo de diversas plataformas de metaversos, lo que permitió la elección de OpenSim como la herramienta para implementar el entorno virtual. OpenSim [22] es un servidor de aplicaciones 3D que facilita la creación de mundos virtuales adaptados a necesidades específicas, en los que la enseñanza juega un papel fundamental. Al ser un software libre y de código abierto, OpenSim permite la creación y el alojamiento propio del EV3D, asegurando un control total sobre el contenido y la facilidad para la ampliación y actualización de escenarios. Además, ofrece interoperabilidad con otros metaversos, como Second Life, lo que posibilita la exportación e importación de inventarios de objetos 3D, ampliando las opciones de interacción de los usuarios.

Para este trabajo, se crearon diferentes escenarios virtuales 3D ad-hoc con el fin de realizar una

experiencia educativa dirigida a personas con discapacidad auditiva. Se buscó determinar si estos entornos y las estrategias de comunicación implementadas constituyen una herramienta eficaz para la enseñanza y el aprendizaje de este público. Los escenarios desarrollados fueron presentados en el XVI Congreso de Tecnología en Educación & Educación en Tecnología, en donde se expuso una demo [23].

A continuación se describe cada uno de los escenarios

Escenario 1: Introducción y Personalización del Avatar

En primer lugar, se diseñó un escenario compuesto por diferentes ambientes, cuyo propósito es introducir a los participantes en el uso del EV3D, permitirles personalizar su avatar y familiarizarse con el entorno. Este escenario incluye dos recorridos para enseñar a los participantes cómo configurar su avatar. En el primer recorrido, los participantes interactúan con carteles guía que permiten visualizar videos explicativos en LSA (Figura 3). El segundo recorrido ofrece opciones para elegir entre avatares predefinidos y contiene instrucciones escritas (Figura 4). Además, los participantes pueden arrastrar y soltar la apariencia deseada desde su inventario.

Escenario 2: Sala de Exposiciones y Debate

En el segundo escenario, se recreó una sala de exposiciones donde se exhiben pósters con información sobre tecnología para personas con discapacidad auditiva (Figura 5). Este espacio también incluye una PC con acceso a un diccionario de vocabulario técnico en informática, desarrollado específicamente para la tesis. Dentro de este escenario, se creó una sala de debate, con mesas redondas y sillas, que fomenta el intercambio de ideas sobre las aplicaciones tecnológicas para la población con discapacidad auditiva (Figura 6). En esta sala se incorpora un tablero de emojis (Figura 7) que los participantes pueden usar durante el debate, permitiéndoles expresar sus emociones. Cada emoji genera una animación en el avatar del participante (Figura 8).

Escenario 3: Aprendizaje Lúdico a través de la Historia del Cine

Un tercer escenario fue creado con el objetivo de presentar un contenido educativo de forma novedosa y lúdica. El tema elegido fue la historia del cine, y el recorrido se realiza en dos etapas dentro de un jardín virtual. En la primera etapa, los participantes deben seguir pistas en pergaminos para aprender sobre la historia del cine mediante videos en LSA. En la segunda etapa, se evalúa el aprendizaje mediante la reconstrucción de una línea de tiempo interactiva (Figura 9), donde los participantes deben ubicar imágenes de eventos históricos en el año correspondiente y responder preguntas de selección múltiple. Los participantes pueden verificar su puntaje al final de la actividad.



Figura 3. Videos en LSA para modificar Apariencia del Avatar (Elaboración Propia)



Figura 4. Apariencias Predefinidas de Avatar (Elaboración Propia)

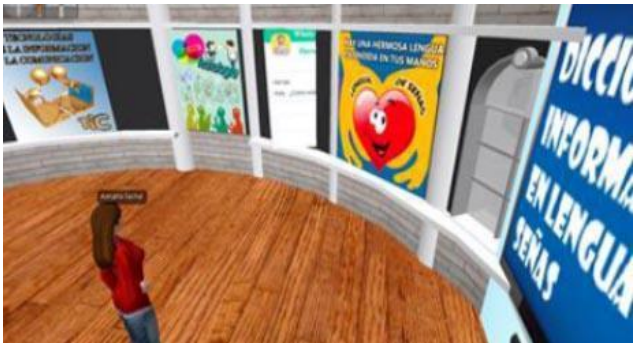


Figura 5. Sala de Exposiciones (Elaboración Propia)



Figura 6. Sala de Debate (Elaboración Propia)

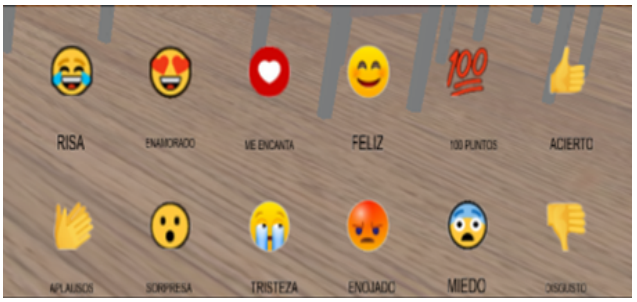


Figura 7. Panel de Emoticones (Elaboración Propia)



Figura 8.Animación de Avatar a partir de la selección de emoticones (Elaboración Propia)



Figura 9.Línea de Tiempo Interactiva (Elaboración Propia)

6. Estudio de Caso

Siguiendo la metodología mencionada en la sección 5.1, se diseñó una experiencia educativa con el EV3D creado con el objetivo de evaluar la participación, motivación y adaptabilidad de personas con discapacidad auditiva en este tipo de entornos. Así la experiencia se estructuró en las etapas propuestas según la metodología (Figura 10).



Figura 10: Etapas del Desarrollo de la Experiencia Educativa

6.1. Convocatoria y Conformación de Grupos Participantes.

Los participantes fueron seleccionados por convocatoria abierta a través de correo electrónico y WhatsApp, conformándose tres grupos:

- Grupo Prueba Piloto: Conformado en diciembre de 2021 por 5 participantes graduados como "Analista de Sistemas" en la institución de nivel terciario ALPI. Este grupo se encargó de realizar pruebas y aportar ante inconvenientes técnicos.
- Grupo de Aprendizaje a Distancia: Compuesto en 2022 por 11 participantes que llevaron adelante la experiencia de manera virtual, tanto sincrónica como asincrónica.
- Grupo del Instituto Integral de Educación (IIDE): Conformado en 2022 por 20 participantes que realizaron toda la experiencia de manera presencial.

6.2. Instalación de software

Para acceder a los entornos virtuales 3D, los participantes instalaron el visor Firestorm Viewer[2]. Antes del primer encuentro,se proporcionó a todos

[2] The Phoenix Firestorm Project http://www.firestormviewer.org

los grupos acceso a un sitio web con videos en LSA realizados por la tesista y material instructivo para guiarlos en la instalación y configuración del software.

Los grupos Prueba Piloto y de Aprendizaje a Distancia instalaron el software de manera autónoma, mientras que acerca del Grupo IIDE se acordó institucionalmente que el proceso de instalación se realizaría en el laboratorio de informática a cargo del técnico del laboratorio y bajo supervisión de la tesista.

6.3. Sesión 1: Funciones Básicas y Modificación de Apariencia del Avatar

El objetivo de esta etapa fue transformar la apariencia del avatar de cada participante, que se presentó en una pizarra digital (Figura 11). La dinámica fue distinta en cada grupo. En el grupo piloto se llevaron a cabo sesiones individuales, en el Grupo a Distancia se utilizaron videoconferencias y en el grupo IIDE se realizaron encuentros presenciales.



Figura 11. Selección de muestra de Avatares logrados

6.4. Sesión 2: Visita a Sala de Tecnología y Debate Grupal

Este encuentro consistió en un recorrido por una sala de tecnología con una muestra de posters referidas a tecnología de utilidad para personas con discapacidad auditiva (Figura 12) y un debate grupal sobre identidad y accesibilidad (Figura 13). Los participantes debatieron sobre la identidad, el uso de la tecnología para personas con discapacidad auditiva y su motivación para utilizar espacios educativos en EV3D. La comunicación se llevó a cabo mediante chat de texto sin uso de voz, recreando un debate presencial, complementado con el uso de tablero de emojis para permitir la expresión de emociones.



Figura 12. Avatares en la sala de exposiciones.



Figura 13. Avatares en la sala de debate

6.5. Sesión 3: Actividades Lúdicas para aprender Historia del Cine

Se implementó una actividad lúdica basada en la historia del cine. Los participantes realizaron una búsqueda del tesoro (Figura 14) y completaron una línea de tiempo interactiva (Figura 15).



Figura 14. Avatares en la búsqueda del tesoro



Figura 15. Avatares armando la línea de tiempo sobre historia del cine

6.6. Evaluación de la experiencia

Al final de cada sesión, los participantes completaron una evaluación mediante un formulario web con un cuestionario diseñado ad-hoc. En particular, el cuestionario utilizado para evaluar la motivación intrínseca fue el Intrinsic Motivation Inventory (IMI)[3], que incluye una escala de valoración del 1 al 7 y 22 ítems distribuidos en diferentes subescalas: 1. Interés/Disfrute; 2. Competencia Percibida; 3. Elección Percibida; y 4. Presión/Tensión .

En el análisis de los resultados, las puntuaciones superiores a 4 en las afirmaciones positivas se consideran como indicativo de una experiencia positiva. Para las afirmaciones negativas, las puntuaciones inferiores a 4 se consideran positivas.

7. Resultados de la Experiencia

En esta sección se presentan los resultados de la evaluación del estudio de caso detallado en la sección 6.

En cuanto a equipamiento se evidencian altas calificaciones en equipos y conexión a Internet como muestra la figura 16.

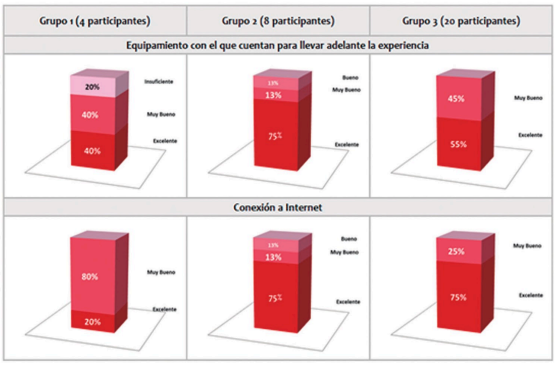


Figura 16: Resultados Equipamiento (Elaboración Propia)

Solo 4 de 36 participantes no pudieron instalar el software debido a la falta de recursos o software obsoleto. Este alto resultado en la instalación se atribuye además a una alta valoración de la calidad del material preparado ad-hoc para la actividad (Figura 17).

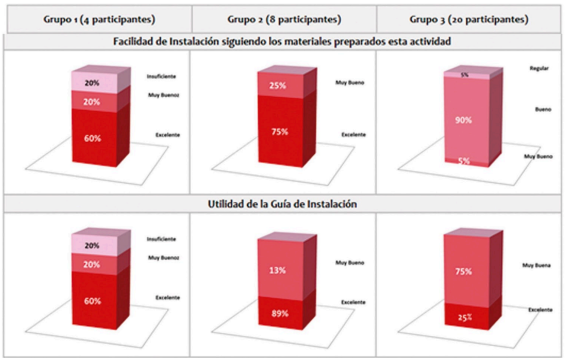


Figura 17: Resultados Instalación de Software (Elaboración Propia)

Todos los participantes lograron cambiar su apariencia, lo que generó una respuesta positiva en cuanto a la motivación, destacando la facilidad del proceso. Los participantes que trabajaron con la apariencia manual valoraron los instructivos paso a paso en LSA dentro del entorno EV3D. Por otro lado, aquellos que utilizaron la función automática de "arrastrar y soltar" expresaron su satisfacción por la simplicidad del proceso.

Además, se observa una evaluación positiva de los participantes sobre el diseño del EV3D y los videos en LSA para modificar el avatar (Figura 18).

[3] IMI (Intrinsic Motivation Inventory https://experienceresearchsociety.org/ux-methods/intrinsic-motivation-inventory-imi/

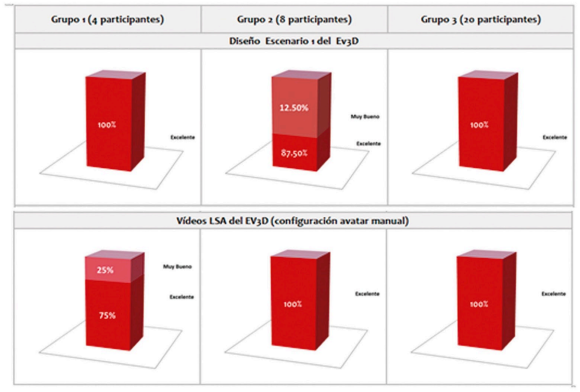


Figura 18: Resultados diseño del EV3D y Vídeos en LSA

En cuanto a la herramienta de chat de texto, los resultados obtenidos muestran una valoración ampliamente positiva tanto en términos de facilidad de uso como en el disfrute del debate (Figura 19). Los participantes destacaron su eficacia para facilitar la interacción y el intercambio de ideas durante la sesión.

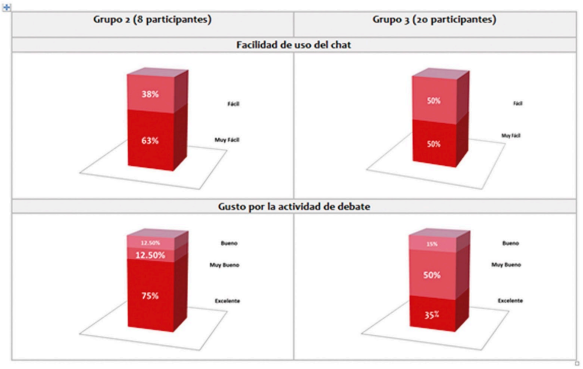


Figura 19: Resultados uso del Chat y Debate

Un aspecto adicional relevante en el ámbito de la comunicación fue el uso del tablero de emojis durante la interacción, un aporte de la tesis que ha sido muy bien valorado. La figura 20 presenta el porcentaje de participantes que utilizaron cada emoji en dos contextos distintos: la sala de exposición y el debate. Se observó un uso predominante del emoji "me encanta" en la sala de exposiciones, con un 75% de los participantes utilizándolo, lo que sugiere una reacción positiva hacia el contenido presentado. Por otro lado, en el contexto del debate, el emoji "ok" fue el más utilizado, con un 71% de los participantes utilizado para señalar acuerdo o comprensión.



Figura 20: Resultados uso de emojis

Además se analizó la relación entre la identidad personal y el avatar utilizado en el EV3D. Se observó que el 57% de los participantes indicó que su avatar representa de manera precisa su identidad, estableciendo una conexión directa con su discapacidad auditiva y por otro lado el 29% de los participantes consideró necesario revelar su discapacidad en el entorno virtual para asegurar la accesibilidad, lo que subraya la importancia de visibilizar esta condición para mejorar la interacción y garantizar un entorno inclusivo (Figura 21).

	Grupo 2	Grupo 3	Totales
Cantidad de participantes	8	20	28
Mujeres Hombres	4 4		
Tu avatar refleja tu identidad	3 (37,5%)	11 (55%)	14 (54%)
Mujeres Hombres	2 (25 %) 1 (12.5%)	4 (20%) 7 (35%)	6 (23%) 8 (31%)
Existe relación entre tu identidad de sordo y tu avatar	4(50%)	12 (60%)	16 (57%)
Mujeres Hombres	2(25%) 2(25%)	3 (15%) 9 (45%)	5 (18%) 11 (39%)
En cualquier EV3D decides revelar tu identidad de sordo	4(50%)	4 (20%)	8 (29%)
Mujeres Hombres	3(37.5%) 1(12,5%)	0 (0%) 4 (20%)	3 (11%) 5 (18%)

Figura 21: Resultados de Identificación del Avatar con Identidad de la persona Sorda

7.1. Resultados del cuestionario IMI

Las tablas 1 y 2 presentan los resultados obtenidos para las subescalas del cuestionario IMI para el Grupo de Experiencia de Aprendizaje a Distancia y el Grupo IIDE, respectivamente. Se evaluaron las dimensiones de interés/diversión, competencia percibida, elección percibida y presión/tensión, proporcionando un análisis detallado de la motivación intrínseca en el uso del EV3D

interés/diversión	6.71
competencia percibida	5.86

elección percibida	6.37
presión/tensión	1.69

Tabla 1: Resultados de cada subescala para el Grupo Experiencia Aprendizaje a Distancia

interés/diversión	6.43
competencia percibida	6.13
elección percibida	6.20
presión/tensión	2.26

Tabla 2: Resultados de cada subescala para el Grupo IIDE

En general, los resultados indican un nivel satisfactorio de motivación, dado que las puntuaciones en interés/diversión, competencia percibida y elección percibida fueron altas, mientras que la subescala de presión/tensión mostró valores bajos. No obstante, se observaron diferencias entre los grupos: el Grupo IIDE obtuvo una mayor percepción de competencia (6.13 frente a 5.86), lo que sugiere que la modalidad presencial ofrece un apoyo más directo y motivacional. A su vez, el Grupo de Experiencia de Aprendizaje a Distancia reportó menores niveles de presión (1.69 frente a 2.26), lo que podría atribuirse a la menor supervisión directa en entornos virtuales.

Estos hallazgos confirman que el EV3D proporciona un entorno adecuado para el aprendizaje inmersivo e interactivo, aunque las diferencias entre modalidades sugieren la necesidad de ajustes para optimizar la experiencia en función del contexto educativo.

8. Conclusiones

Este estudio ha investigado el potencial de los EV3D en la educación de personas con discapacidad auditiva, explorando su capacidad para mejorar la comunicación, la socialización y la motivación intrínseca de los participantes. A lo largo de la investigación se abordaron diversos aspectos teóricos y metodológicos que derivaron en las contribuciones de esta tesis.

Uno de los principales aportes de esta investigación ha sido el desarrollo de 11 guías de diseño para la creación de EV3D educativos dirigidos a personas con discapacidad auditiva. Estas guías abarcan aspectos pedagógicos, lingüísticos y tecnológicos, proporcionando recomendaciones para facilitar la planificación, el diseño y la implementación de experiencias educativas en EV3D. Además, se propuso una metodología en cinco etapas que organiza el proceso de desarrollo de actividades educativas, con el fin de alcanzar la motivación, el aprendizaje y la comunicación de los participantes.

A través de la implementación de estos conceptos, se desarrolló un entorno virtual 3D en OpenSim que fue evaluado positivamente por los estudiantes. La experiencia educativa ofreció diversas herramientas de comunicación, como el chat textual y un tablero de emojis, orientado a enriquecer la expresión de emociones y una mejor interacción entre los participantes.

Los resultados del estudio de caso mostraron que los estudiantes, tanto en modalidad presencial como online, experimentaron una alta motivación intrínseca. Las subescalas del cuestionario IMI revelaron una actitud positiva hacia el uso del EV3D, destacando la percepción de competencia y la autonomía en la elección de las actividades. Sin embargo, se observaron diferencias significativas en la percepción de competencia, siendo los grupos presenciales los que mostraron una mayor sensación de dominio de la herramienta, lo que podría estar relacionado con factores como la edad y el contexto de la participación.

En términos metodológicos, se observó que la flexibilidad de la metodología permitió una integración exitosa de la tecnología en diversos contextos, superando las barreras tecnológicas iniciales y promoviendo una alta participación de los estudiantes. La experiencia educativa diseñada también evidenció el valor de la gamificación y de una planificación pedagógica detallada para mantener el interés y la motivación de los participantes. El chat textual, el tablero de emojis y la cartelería dentro del EV3D con videos en LSA fueron altamente valorados por los usuarios, evidenciando su efectividad para superar las barreras de comunicación que enfrentan las personas con discapacidad auditiva.

En conclusión, los hallazgos de esta investigación no solo demuestran el potencial de los EV3D como herramientas educativas inclusivas, sino que

también proporcionan un marco práctico y accesible para diseñar experiencias de aprendizaje para personas con discapacidad auditiva. Los resultados obtenidos abren nuevas posibilidades para el uso de entornos virtuales 3D en la educación inclusiva, contribuyendo significativamente a la investigación y el desarrollo de nuevas prácticas pedagógicas adaptadas a la diversidad de necesidades de los estudiantes.

9. Líneas de Investigación de Trabajo Futuras

A partir de los hallazgos de esta investigación, se identifican diversas direcciones para futuras exploraciones en el campo de los EV3D para personas con discapacidad auditiva. En primer lugar, se plantea el desarrollo de avatares señantes, incorporando animaciones en Lengua de Señas Argentina (LSA) para enriquecer la comunicación. Además, se propone la expansión y aplicación de las guías de diseño en distintos entornos educativos, evaluando su adaptabilidad a diversos perfiles de usuarios. Asimismo, se considera la ampliación del tablero de emojis, incorporando nuevas emociones y animaciones para mejorar la expresividad en la interacción. Finalmente, se plantea la integración de personas oyentes y con discapacidad auditiva en experiencias compartidas dentro de EV3D, analizando la dinámica de interacción en entornos inclusivos.

10. Referencias

[1] Cuadrillero, O. (2012). Atendiendo Necesidades. La Educación para las personas sordas o con deficiencia auditiva. Obtenido de http://atendiendonecesidades.blogspot.com.ar/2012/11/la-educacion-para-las-personas-sordas-o.html

[2] Zappalá, D., Köppel, A., & Suchodolski, M. (2010). Inclusión de TIC en escuelas para alumnos sordos e hipoacúsicos. Buenos Aires, Argentina: Conectar Igualdad.

[3] Zappalá, D., Köppel, A., & Suchodolski, M. (2011). Serie Computadoras Portátiles para las escuelas de educación especial. Inclusión de TIC en escuelas para alumnos sordos e hipoacúsicos. Córdoba, Argentina: Conectar Igualdad.

[4] Baró Miró, M. (2023). Uso del metaverso en la enseñanza de ELE: propuesta de aplicación para el aprendizaje del léxico. UNIVERSITAT POLITÈCNICA DE VALÈNCIA.

[5] Adell, F. (28 de Noviembre de 2012). Mundos Virtuales y Entornos Educativos Complejos. Obtenido de http://mosaic.uoc.edu/2012/11/28/mundos-virtuales-y-entornos-educativos-complejos/

[6] Aguilar Martínez, J. L., Alonso López, M., Arriaza Mayas, J. C., Brea San Nicólas, M., Cairón Ceballos, M. I., Camacho Hermoso, M. D., y otros. (2012). Manual de Atención al Alumnado con Necesidades Específicas de Apoyo Educativo Derivadas de Discapacidad Auditiva. Andalucía: Tecnographic, S.L.

[7] Arco, J. L., & Fernández Castillo, A. (2004). Manual de Evaluación e Intervención Psicológica en Necesidades Educativas Especiales. Madrid, España: McGRAW-HILL/INTERAMERICANA DE ESPAÑA, S. A. U.

[8] Smith, K. (2010). The use of virtual worlds among people with disabilities. In Proceedings of the International Conference on Universal Technologies, (págs. pp. 15-21). Oslo, Norway

[9] López Falcón, G. (2014). Producción del curso virtual sobre el uso del software "OpenSim" como herramienta de apoyo didáctico. Revista de Lenguas Modernas, N° 21, 2014 / ISSN: 1659-1933, 347-365.

[10] Warburton, S. (2008). Muves and second lives: exploring education in virtual worlds. Excellence in Teaching Conference 2008 Annual Proceedings, (págs. pp. 119–127).

[11] Baños González, M., Rodríguez García, T., & Rajas Fernández, M. (2014). Mundos virtuales 3D para la comunicación e interacción en el momento educativo on-line. Historia y Comunicación Social, 19(Nº Esp. Enero), 417-430.

[12] Solis, M. Á., & Prats, E. C. (2014). OpenSIM: Oportunidades de comunicación e interacción en mundos virtuales accesibles. In

Actas del VI Congreso Internacional sobre Aplicación de Tecnologías de la Información y Comunicaciones Avanzadas (ATICA 2014), (págs. pp. 189-196). Universidad de Alcalá de Henares (España)

[13] Tosete, F. (2009). Para qué pueden servir un mundo virtual como Second Life y la realidad aumentada. Anuario ThinkEPI, (págs. 185-192).

[14] Padovani, F. G. (2013). Utilización de OpenSim para la Educación. Buenos Aires, Argentina: Escuela de Tecnología.

[15] Fernández Sánchez, N. (2014). OEA - Portal Educativo. Recuperado el 2023, de Entornos Virtuales de Aprendizaje. Instaurando un Mundo Virtual portable con OpenSim autónomo, en 24 horas: https://recursos.educoas.org/publicaciones/entornos-virtuales-de-aprendizaje-instaurando-un-mundo-virtual-portable-con-opensim

[16] OpenSimulator. (2022). OpenSimulator. Recuperado el 2023, de OpenSimulator: http://opensimulator.org/wiki/Main_Page

[17] Díaz Gandasegui, V. (2013). Entornos virtuales para el desarrollo de la educación inclusiva: Una mirada hacia el futuro desde el pasado de Second Life.

[18] Ramón, H. D., Russo, C. C., Sarobe, M., Alonso, N., Esnaola, L., Ahmad, T., y otros. (2014). El uso de los Entornos Virtuales 3D como una herramienta innovadora en propuestas educativas mediadas con tecnología. Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología, 12, 72-80.

[19] Contreras Masedo, D., Fernández-Avilés Pedraza, D., & Salazar Calderón, J. C. (2023). Universidad Politécnica de Madrid. Recuperado el 2023, de Meta-manual Laboratorios 3D: https://3dlabs.upm.es/web/manuales/Meta_Manual_Laboratorios_3D.pdf

[20] Carr, D. (Marzo de 2010). Constructing Disability in Online Worlds; Conceptualising Disability in Online Research' London Review of Education: Special Issue 'Being Online: A Critical View of Identity and Subjectivity in New Virtual Learning Spaces' 8. 1. Obtenido de https://learningfromsocialworlds.wordpress.com/9a-deaf-in-second-life-paper/

[21] Ferreira Szpiniak, A. (2013). Diseño de un modelo de evaluación de entornos virtuales de enseñanza y aprendizaje basado en la usabilidad. La Plata - Buenos Aires: Facultad de Informática - Universidad Nacional de La Plata.

[22] OpenSimulator. [Online]. Available: http://opensimulator.org/wiki/Main_Page. [Accessed 10, 2023]

[23] A.S. Fachal, M.J. Abásolo y C.V. Sanz, CV. Entorno Virtual 3D en OpenSim para el trabajo con estudiantes con discapacidad auditiva, presentada en XVI Congreso de Tecnología en Educación & Educación en Tecnología, Universidad Nacional de La Plata, Buenos Aires, 10 y 11 de junio de 2021.

Información de Contacto de los Autores:

Adriana Silvia Fachal
50 y 120 2do Piso
La Plata - Argentina
afachal@hotmail.com

María José Abásolo
50 y 120 2do Piso
La Plata - Argentina
mjabasolo@lidi.info.unlp.edu.ar

Cecilia V. Sanz
50 y 120 2do Piso
La Plata - Argentina
csanz@lidi.info.unlp.edu.ar

Fachal Adriana Doctora en Ciencias Informáticas, UNLP, Argentina. Magíster en Ingeniería del Software, Instituto Tecnológico de Buenos Aires (ITBA), Argentina.

Abásolo María José Doctora en Informática (2002, Departamento de Ciencias Matemáticas e Informática de la Universidad de las Islas Baleares). Investigadora de la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires. Profesor Asociado de la Facultad de Informática de la Universidad Nacional de La Plata (UNLP), La Plata, Argentina. Investigadora miembro del III-LIDI. Directora de la Especialización en Computación Gráfica, Imágenes y Visión por Computadora

Sanz Cecilia V. Doctora en Ciencias de la Facultad de Ciencias Exactas de la Universidad Nacional de La Plata, Argentina. Profesora Titular DE de la Facultad de Informática de la UNLP. Investigadora del III LIDI - CIC. Directora de la Maestría en Tecnología Informática aplicada en Educación. Facultad de Informática. UNLP

Reconstrucción antropométrica 3D de bajo costo basada en procesamiento de imágenes y Deep Learning

Dra. Magda Alexandra Trujillo Jiménez [1, 2]
mtrujillo@cenpat-conicet.gob.ar

Directores de Tesis: Dr. Claudio Delrieux [1], Dr. Rolando Gonzalez-José [2]
Universidad y Fecha de exposición: Universidad Nacional del Sur, 14 de junio de 2024

[1] Instituto de Ciencias e Ingeniería de la Computación ICIC, Universidad Nacional del Sur
[2] Instituto Patagónico de Ciencias Sociales y Humanas IPCSH, CCT-CONICET-CENPAT

Introducción

La antropometría, el estudio de las medidas y formas del cuerpo humano, es esencial en disciplinas como la salud, la nutrición y la ergonomía. Sin embargo, los métodos tradicionales presentan importantes limitaciones: requieren instrumentos costosos y especializados, personal entrenado y están sujetos a errores humanos, lo que dificulta su aplicación en estudios poblacionales a gran escala. Esta tesis busca democratizar el acceso a la antropometría 3D mediante el desarrollo de un sistema de reconstrucción del cuerpo humano que sea preciso, accesible y eficiente, utilizando dispositivos móviles como herramientas de captura. En la intersección entre la informática y la antropometría, esta investigación propone un enfoque innovador basado en el procesamiento avanzado de imágenes y la aplicación de algoritmos de Machine Learning y Deep Learning para el escaneo y análisis de la forma humana. Al superar las limitaciones de los métodos convencionales, este sistema de bajo costo permite obtener modelos 3D confiables, medibles y de alta calidad. Además, la integración de datos antropométricos con variables como la salud, los hábitos alimentarios y el estilo de vida abre nuevas oportunidades en investigación científica, prevención de enfermedades y aplicaciones en sectores como la medicina y el diseño.

Objetivos y Metodología

El objetivo principal de esta tesis es investigar y desarrollar algoritmos que permitan la reconstrucción antropométrica 3D utilizando dispositivos móviles. Los objetivos específicos incluyen:

- Desarrollar métodos de detección, segmentación y reconstrucción 3D basados en Deep Learning y Structure from Motion (SfM).
- Crear prototipos que permitan la reconstrucción 3D y la identificación de landmarks anatómicos.
- Buscar y correlacionar información poblacional para inferir tendencias antropológicas y sanitarias.

La tesis se divide en cuatro partes principales:

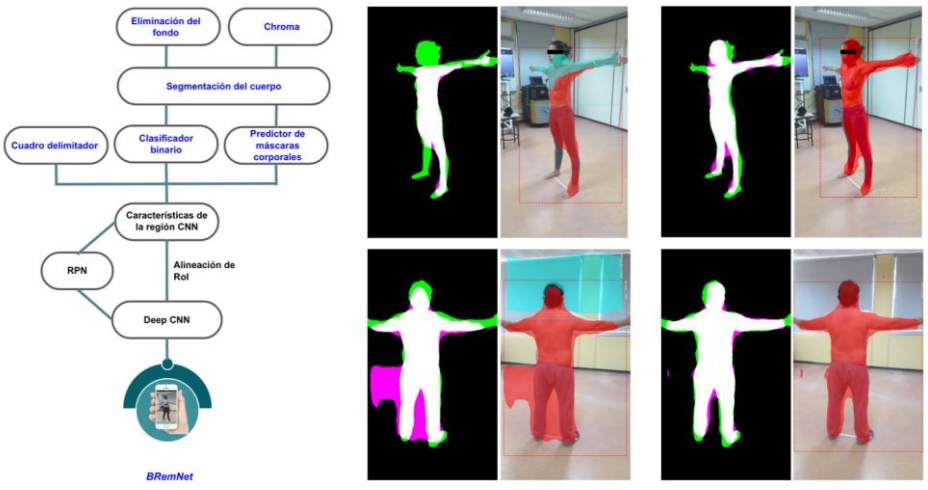
1. **Introducción:** Se detallan los objetivos, aplicaciones y contribuciones académicas de la tesis.

2. **Metodologías:** Se presentan los métodos utilizados para la reconstrucción 3D, incluyendo el procesamiento de imágenes, la segmentación corporal, la reconstrucción de nubes de puntos 3D y la adquisición de mallas 3D, todo ello con un fuerte componente de Deep Learning.
3. **Aplicaciones:** Se exponen los conceptos clave de la antropometría y los tres casos de estudio en los que se aplicó la metodología desarrollada.
4. **Conclusiones:** Se presentan las conclusiones de la tesis y se discuten los trabajos en curso y futuros.

Contribuciones

Esta tesis presenta varias contribuciones significativas al campo de la reconstrucción antropométrica 3D:

- **BRemNet: Desarrollo de un modelo de segmentación de cuerpos humanos en imágenes y video.** Una segmentación precisa es esencial para eliminar el ruido de fondo y garantizar que solo la figura humana se emplee en la reconstrucción 3D. Este modelo, basado en Redes Neuronales Convolucionales, ha demostrado una alta eficiencia en la clasificación y segmentación de la figura humana en entornos complejos, incluso en imágenes en movimiento e incompletas. BRemNet emplea un clasificador binario que se entrena utilizando pesos pre-entrenados del conjunto de datos MS COCO. El entrenamiento se lleva a cabo con un conjunto de datos específico de 200 frames que capturan diversos cuerpos en diferentes ubicaciones del dataset propio de esta tesis.



(a) (b)

Figura 1: (a) Arquitectura de BRemNet. (b) Ejemplos de segmentación: Mask R-CNN (izquierda) y BRemNet (derecha). La máscara segmentada se superpone en rojo a los frames reales. En Mask R-CNN las máscaras de otros objetos identificados se superponen en cian. En las máscaras, los verdaderos positivos en blanco, los falsos positivos en magenta y los falsos negativos en verde.

Se utilizaron cuatro métricas de calidad para evaluar la segmentación: la métrica de pérdida de Hamming, el índice de Jaccard, la medida F1 y la precisión. BRemNet muestra mejores resultados que Mask R-CNN en todas las métricas de calidad y presenta mayor consistencia

en la segmentación de los *frames* de video, especialmente en videos con ruido o movimiento irregular (ver Figura 1). El código está alojado en: https://github.com/aletrujim/BRemNet.

- **Deep Learning aplicado a la obtención de landmarks corporales y segmentación corporal.** Para la detección de *landmarks*, se implementaron y compararon modelos como OpenPose y BlazePose, utilizando *frames* de video segmentados por BRemNet como entrada (ver Figura 2a). Estos modelos, basados principalmente en redes CNNs, procesan las imágenes para extraer características relevantes como bordes y formas, y luego utilizan estos datos para identificar los *landmarks* anatómicos en la imagen, principalmente articulaciones, estimando sus correspondientes posiciones y orientaciones. Cabe destacar que OpenPose utiliza Visual Geometry Group (VGG19) en sus primeras capas para crear mapas de características para la imagen de entrada. Por otro lado, BlazePose, desarrollado por Google e implementado en MediaPipe, utiliza 33 *landmarks* corporales, incluyendo algunos *landmarks* faciales, y en las manos y pies. Este modelo cuenta con dos fases, la primera es un detector que localiza la región de interés (RoI) de la silueta de la persona dentro del *frame* de video y la segunda, es un rastreador que predice los 33 *landmarks* a partir de esa RoI. El código está alojado en: https://github.com/aletrujim/body-landmarks.

  En cuanto a la segmentación corporal, se adaptó una versión de la Fully Convolutional Networks (FCN) para segmentar las máscaras binarias resultantes de BRemNet en regiones como cabeza, torso, brazos y piernas (ver Figura 2b). Esta FCN cuenta con capas de convolución (Conv), agrupación (Pool) y deconvolución (Deconv), y aplica capas de convolución traspuesta, fusionadas con las salidas de las capas inferiores para generar una salida de clasificación y segmentación más fina. El código está alojado en: https://github.com/aletrujim/body-segmentation.

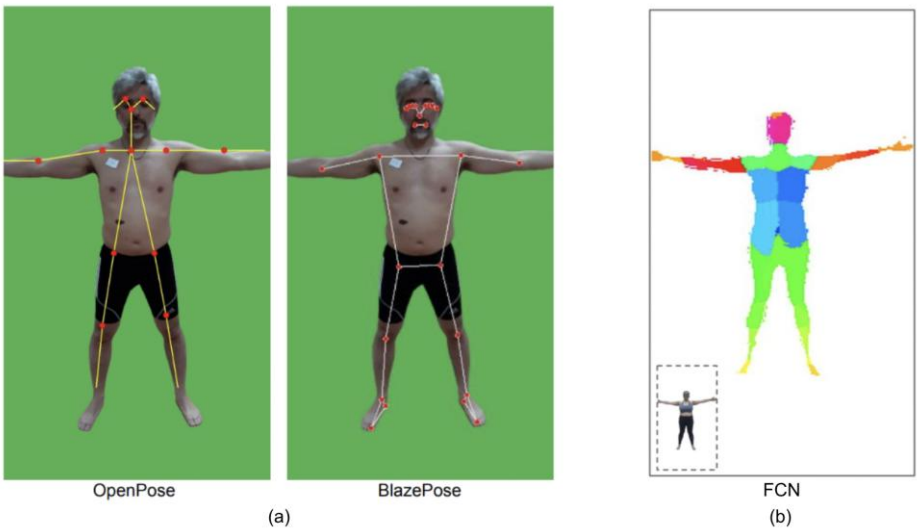


Figura 2: (a) Comparación de resultados entre OpenPose y BlazePose. (b) Ejemplo de segmentación corporal con la red FCN. Para fines ilustrativos, se muestra el frame segmentado y no la máscara binaria (cuadro inferior derecho).

- **body2vec: Reconstrucción 3D a partir de videos.** Se desarrolló un método para generar nubes de puntos 3D a partir de videos utilizando SfM (ver Figura 3). El método body2vec es una técnica desarrollada para la reconstrucción 3D a partir de videos, que se distingue por la

utilización de fotogramas previamente segmentados por BRemNet. Este enfoque permite generar nubes de puntos 3D más limpias y precisas, lo que mejora significativamente la calidad de la reconstrucción. La segmentación previa con BRemNet juega un papel crucial, ya que al segmentar los fotogramas de video, se eliminan los puntos espurios del fondo, lo que reduce el ruido y mejora la calidad de la nube de puntos resultante. Esto permite obtener reconstrucciones corporales en forma de nubes de puntos 3D aptas para hacer antropometría. Para evaluar la calidad de las nubes de puntos, se comparan las generadas a partir de fotogramas de video sin procesar y fotogramas segmentados con BRemNet, así como con mallas 3D obtenidas con un escáner LiDAR. En promedio, las nubes de puntos generadas con los fotogramas segmentados por BRemNet están compuestas por 37.265 puntos. Se utilizan escáneres LiDAR para obtener mallas 3D corporales, usadas como referencia para la validación de los métodos desarrollados con mediciones de perímetros abdominales como cintura y cadera (ver Figura 4). Se aplican métodos de post-procesamiento como el suavizado Laplaciano y el cierre de agujeros.

- **Evaluación rigurosa de las nubes de puntos:** Se llevó a cabo una evaluación métrica rigurosa de las nubes de puntos generadas mediante SfM, comparándolas con mallas 3D obtenidas con un escáner LiDAR, utilizando el método ICP (punto más cercano iterativo). El objetivo fue validar cuantitativamente la mejora en la calidad de las reconstrucciones 3D al utilizar la segmentación con BRemNet. Se usó ICP para alinear las nubes de puntos y las mallas 3D de referencia, minimizando la distancia entre ellas de forma iterativa. Se evaluaron métricas de precisión como el error cuadrático medio (RMSE), la distancia media y la desviación estándar, para cuantificar la exactitud de las reconstrucciones 3D. Los resultados demostraron que la segmentación con BRemNet mejora significativamente la exactitud en las reconstrucciones 3D. El RMSE medio de la nube de puntos sin procesar fue de 12,11 cm, mientras que el error se redujo sustancialmente a 2,02 cm en la nube limpia, lo que indica una mejora sustancial en la precisión gracias a BRemNet. En resumen, esta evaluación validó la efectividad de la segmentación con BRemNet para mejorar la calidad de las reconstrucciones 3D.
- **Análisis de descriptores de forma:** Se realiza un análisis exhaustivo de descriptores de forma para las nubes de puntos reconstruidas, abarcando características geométricas esenciales. Entre ellos, la curvatura mide la desviación de la superficie respecto a un plano, la linealidad evalúa la similitud con una línea recta, la omnivarianza cuantifica la variación en todas las direcciones, la planaridad determina la semejanza con una superficie plana y la esfericidad mide la similitud con una forma esférica. Estos descriptores proporcionan información clave para analizar y procesar la morfología humana. Además, su cálculo facilita la cuantificación de características corporales y contribuye a la limpieza de puntos espurios en las nubes de puntos reconstruidas.

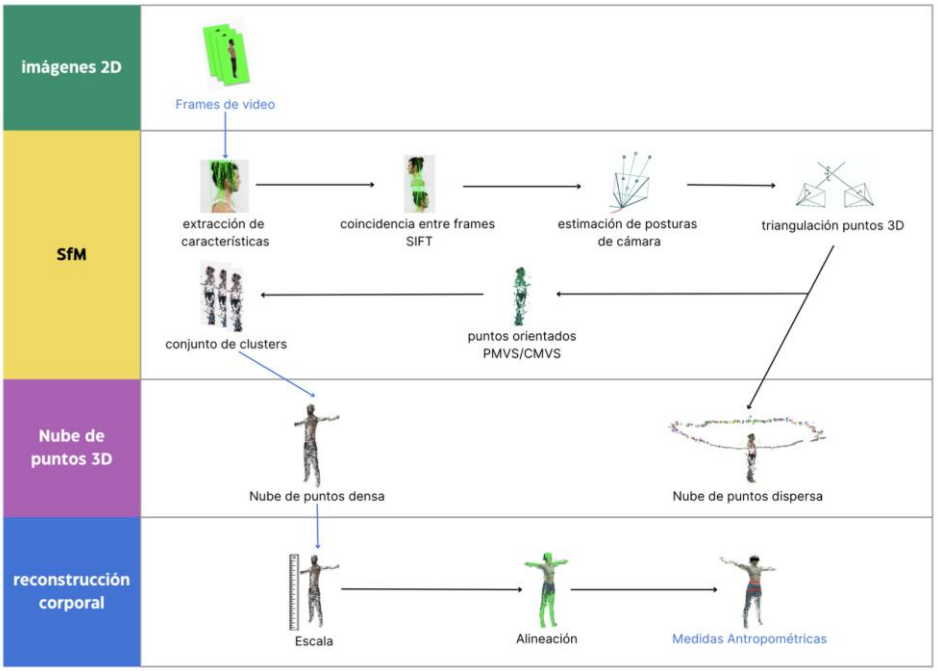


Figura 3: Esquema del método body2vec para crear nubes de puntos 3D que permitan la toma de medidas antropométricas.

Mediante el uso de filtros basados en estos descriptores, se pueden identificar y eliminar los "outliers" o valores atípicos que no representan fielmente la forma del cuerpo. Por ejemplo, el filtro "Radius Outlier Removal" elimina los puntos que tienen pocos vecinos en un radio específico. Este proceso de filtrado es esencial para mejorar la calidad y precisión de las nubes de puntos, asegurando que los análisis y mediciones posteriores se basen en datos confiables y representativos. De esta manera, se logra una representación más precisa y depurada de la forma corporal, optimizando los resultados de los análisis antropométricos y morfométricos.

- **Aplicación de métodos de reconstrucción de superficies:** Se exploran diferentes técnicas para reconstruir superficies a partir de las nubes de puntos. Se evalúan métodos tradicionales de visión computacional como Ball Pivoting y Screened Poisson para generar mallas a partir de nubes de puntos, aunque los resultados no fueron del todo satisfactorios. Como alternativa, se investigaron métodos basados en autoencoders, como FoldingNet, para lograr una reconstrucción más precisa.
- **Implementación de Deep Learning para la reconstrucción de mallas:** Se explora el uso de redes neuronales profundas para el procesamiento de mallas 3D. Se probó el modelo PIFuHD, que utiliza una función implícita alineada por píxeles para generar mallas 3D de alta resolución. Este modelo se basa en un codificador de imágenes totalmente convolucional y una función implícita continua representada por una red neuronal de perceptrón multicapa (MLP). La red aprende una función implícita sobre el espacio 3D con características de imagen alineadas con píxeles, lo que permite preservar los detalles locales presentes en la imagen. Su arquitectura consta de dos niveles de módulos PIFu. El primero, conocido como módulo grueso, toma como entrada imágenes de baja resolución (512×512) y se centra en integrar la información geométrica global.

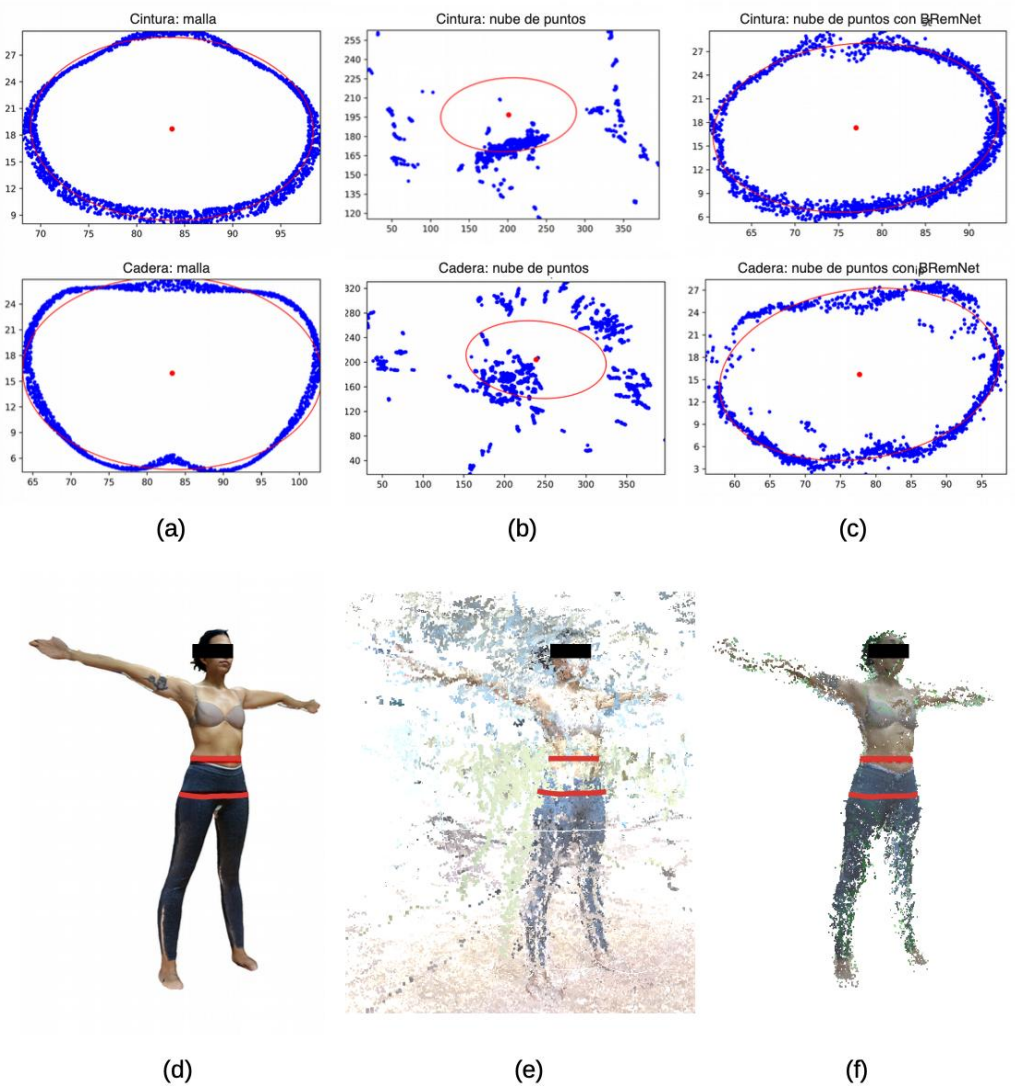
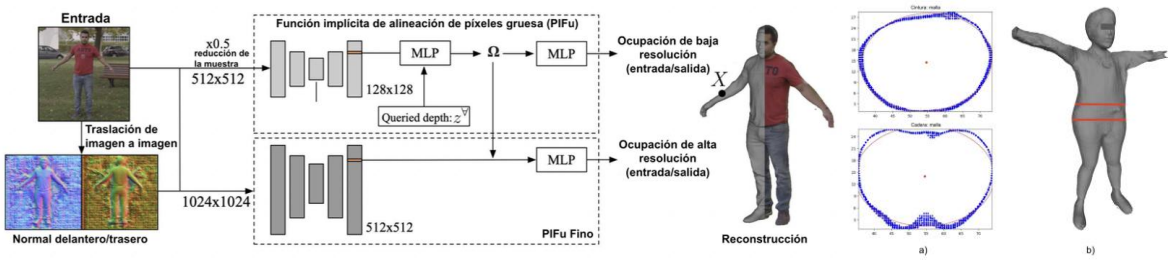


Figura 4: Las tres reconstrucciones 3D de cuerpo completo: (a) malla obtenida con escáner LiDAR, (b) nube de puntos basada en SfM a partir de vídeo sin procesar, y (c) nube de puntos a partir de vídeo limpio filtrado por BRemNet.



(a) (b)

Figura 5: (a) Arquitectura de PIFuHD basada en 2 niveles de entrenamiento. (b) Representación de las mediciones cintura y cadera en 3D de PIFuHD. Figuras obtenidas de cintura y cadera. Malla generada por el modelo.

El segundo nivel, o módulo fino, utiliza características de incrustación 3D extraídas del módulo grueso e imágenes de mayor resolución (1024×1024). Si bien los modelos 3D

resultantes de PIFuHD son aceptables dada su flexibilidad en cuanto a requerimientos de hardware y tiempos de procesamiento, la calidad tanto visual como en medidas es un desafío a mejorar (ver Figura 5). Además, algunos modelos no logran reconstruir ciertas partes del cuerpo. Incorporar Deep Learning y técnicas avanzadas computacionales, permiten optimizar la reconstrucción y generar representaciones más precisas y realistas del cuerpo humano. Estas innovaciones no solo mejoran la calidad visual y la fidelidad de las medidas, sino que también amplían las posibilidades de aplicación en diversos campos, desde la salud hasta la industria del diseño y la ergonomía. El código está alojado en: https://github.com/aletrujim/3Dtest/blob/main/PIFu_test.ipynb.

Aplicaciones

- **Antropometría**: La tesis revisa la antropometría clásica y las innovaciones de la antropometría 3D. La antropometría clásica ha sido clave para estudiar la diversidad humana mediante mediciones como estatura, peso y circunferencias, permitiendo analizar diferencias poblacionales y su relación con factores genéticos y ambientales. Por otro lado, la antropometría 3D ha transformado el estudio de la morfología humana con tecnologías de escaneo y reconstrucción tridimensional, logrando mediciones más precisas y detalladas. Este enfoque permite generar modelos digitales realistas, con aplicaciones en medicina, diseño industrial, moda y cirugía reconstructiva, además de facilitar el monitoreo de pacientes con obesidad. Al integrar estas innovaciones, la antropometría 3D amplía las posibilidades de análisis, proporcionando una visión más completa de la variabilidad humana en distintos campos científicos y tecnológicos.
- **Protocolo de toma de datos con dispositivos móviles:** Se desarrolló un protocolo sencillo y no invasivo para la toma de datos con dispositivos móviles, que incluye instrucciones detalladas para la captura de videos y escaneos, y la obtención de medidas antropométricas. Este protocolo fue elevado como base del protocolo del Programa de Referencia y Biobanco Genómico de la Población Argentina PoblAr. Además, su diseño permite la recolección de datos en entornos diversos sin necesidad de equipamiento especializado, facilitando su implementación en estudios poblacionales a gran escala. Su validación en diferentes contextos ha demostrado su eficacia y reproducibilidad, garantizando la calidad y precisión de los modelos 3D obtenidos.
- **Casos de estudio:** La tesis tiene como objetivo desarrollar métodos para obtener mediciones antropométricas precisas a partir de modelos 3D reconstruidos, permitiendo un análisis morfométrico detallado del cuerpo humano. Para garantizar la generalización y aplicabilidad de estos métodos, su validación se llevó a cabo en tres muestras poblacionales distintas, demostrando su robustez y utilidad en diversos contextos. A su vez, se explora la relación entre los datos antropométricos 3D y variables de salud y hábitos alimentarios. Se especifican mediciones clave, como altura, peso y circunferencias de cintura y cadera, analizándolas en relación con indicadores esenciales del estado nutricional y riesgo metabólico. El IMC que permite clasificar el peso corporal, el ICC evalúa la distribución de grasa y su relación con enfermedades cardiovasculares, y el ICT es clave para detectar obesidad abdominal. Estos parámetros, obtenidos a partir de las medidas de los modelos 3D, ofrecen una herramienta precisa y accesible para estudios poblacionales y monitoreo de la salud. Finalmente, se evalúa

el error de las mediciones en comparación con datos antropométricos reales, destacando la mejora significativa en la precisión de la reconstrucción 3D mediante el uso de BRemNet.

- **Proyecto Raíces y Patagonia 3D Lab:** Este estudio se llevó a cabo en Puerto Madryn, Argentina, con una muestra de 299 personas pertenecientes a una población cosmopolita y urbana. El objetivo del proyecto es estudiar los diferentes factores genéticos que pueden influir en la diversidad fenotípica y de qué manera esta diversidad es autopercibida por la sociedad. Para ello, se generaron muestras de referencia que combinan una gran cantidad de datos, como imágenes faciales y corporales en 3D, antropometría (ver Figura 6), incluidos valores obtenidos por bioimpedancia, datos genéticos obtenidos posteriormente a partir de una muestra de sangre, datos socioculturales de hábitos y estilo de vida asociados al sedentarismo, tabaquismo, dieta, y factores económicos, un cuestionario sobre antecedentes de salud familiares e individuales, y a la ancestría y autopercepción de la misma, todas variables involucradas en la compleja relación entre salud, nutrición y sociedad.
- **ECHA (Emoción, Cognición y Hábitos Alimentarios):** Realizado en Bahía Blanca, Argentina, este estudio analizó la relación entre factores psicológicos, hábitos alimentarios y características antropométricas en 106 personas de 18 a 60 años. Se integraron datos de reconstrucción 3D con mediciones tradicionales para evaluar la utilidad de estos modelos en estudios interdisciplinarios sobre nutrición y salud. Se exploró cómo variables emocionales y cognitivas, como el estrés o la percepción corporal, influyen en la composición y forma del cuerpo (ver Figura 7).

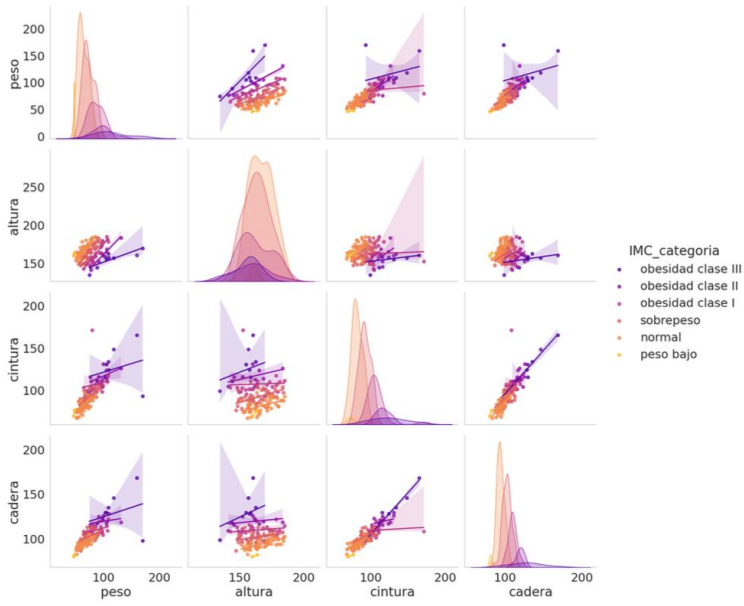


Figura 6: Raíces y Patagonia 3D Lab: Medidas antropométricas. Categorías según el índice de masa corporal.

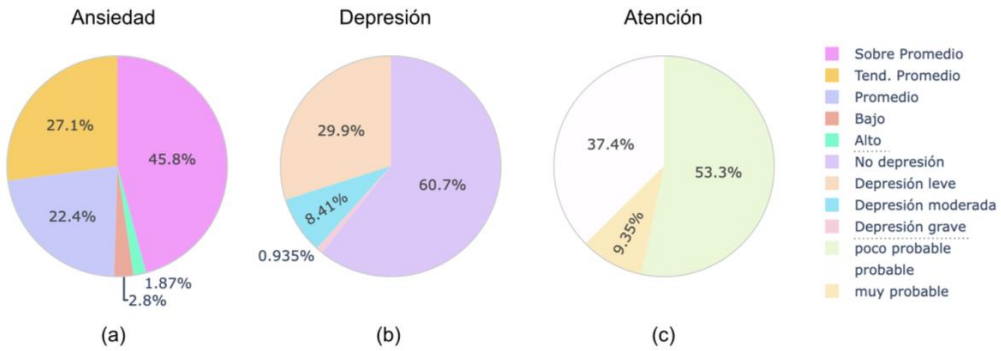


Figura 7: ECHA: algunas variables que describen la muestra. (a) Ansiedad, (b) depresión y (c) atención.

- **SER (Society, Environment and Reproduction Research):** Este estudio se llevó a cabo en una población Maya Kaqchiquel en Guatemala, centrándose en 116 mujeres en edad reproductiva. Se investigó la relación entre factores socioculturales, ambientales y reproductivos con la morfología corporal, utilizando técnicas de escaneo 3D para documentar variaciones antropométricas en una comunidad con características biológicas y culturales únicas. La recolección de datos permitió generar modelos tridimensionales de alta precisión, contribuyendo a estudios sobre la relación entre salud materna, nutrición y estructura corporal en esta población indígena.

Estos tres estudios complementan el análisis de la tesis al abordar la reconstrucción antropométrica 3D en poblaciones con diferentes características socioeconómicas, culturales y ambientales, lo que permite validar la metodología propuesta en diversos contextos.

- Se tratan los aspectos éticos relevantes a la investigación, asegurando la protección de los derechos y el bienestar de los participantes. Se menciona el uso de consentimiento informado.

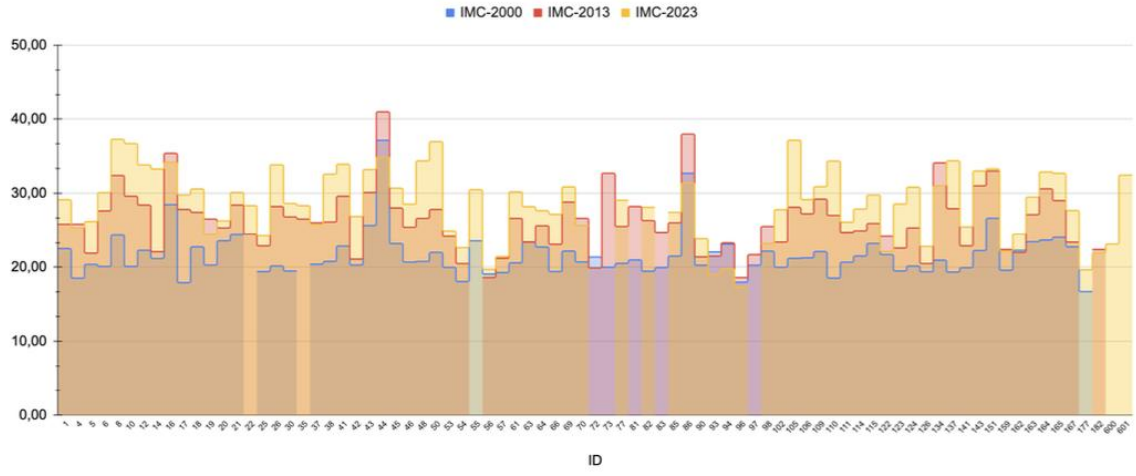


Figura 8: SER: comparación del IMC en la muestra. Cálculo del IMC en el año 2000 (en azul), 2013 (en rojo) y 2023 (en amarillo).

Conclusiones

Esta tesis doctoral representa un avance significativo en la reconstrucción antropométrica 3D y su análisis morfométrico, ofreciendo una alternativa precisa, accesible y eficiente a los métodos

tradicionales. Los métodos desarrollados, junto con el análisis de datos en diversas poblaciones, brindan herramientas valiosas para diversas disciplinas científicas y aplicaciones prácticas. La integración del procesamiento de imágenes, el aprendizaje automático y el aprendizaje profundo ha resultado eficaz para superar los desafíos existentes en este campo.

En el contexto específico de la informática, esta tesis doctoral destaca por su aplicación innovadora de técnicas de Deep Learning y visión computacional para resolver un problema complejo en el ámbito de la antropometría. La implementación de algoritmos de segmentación semántica y por instancias, la reconstrucción de nubes de puntos 3D a partir de fotogrametría y la mejora de mallas 3D mediante Deep Learning demuestran un enfoque original y efectivo para la creación de modelos antropométricos precisos y detallados. Asimismo, la validación de los métodos desarrollados en diversas poblaciones garantiza su aplicabilidad y robustez en diferentes contextos, lo que representa un valor añadido para la comunidad informática. Esta investigación contribuye al avance del conocimiento en el campo de la informática al proponer soluciones innovadoras para el análisis de datos tridimensionales y su aplicación en la reconstrucción antropométrica.

Trabajo en curso y futuro

En el marco del trabajo en curso, la investigación se centra en ampliar y profundizar los resultados obtenidos, analizando y correlacionando los datos recolectados en diversos proyectos con factores de salud, hábitos alimentarios, ansiedad y aspectos somáticos de la maduración reproductiva y el envejecimiento. Se están evaluando aspectos relacionados con enfermedades crónicas, hábitos alimentarios y ansiedad, así como la maduración reproductiva en diversas poblaciones. Además, se están explorando datos y modelos basados en Deep Learning para obtener un descriptor de forma corporal que permita describir la forma humana obtenida en 3D con una matriz más pequeña, similar a un enfoque estilo PCA. Este descriptor facilitará el análisis y la comparación de la forma corporal en diferentes contextos, buscando simplificar la representación de datos antropométricos.

En cuanto al trabajo futuro, se plantea seguir mejorando y optimizando los algoritmos de segmentación y reconstrucción 3D, con un enfoque en la precisión y la eficiencia. Se espera expandir estos métodos a campos interdisciplinarios como la medicina, la moda y la ciencia forense, explorando enfoques de automatización y Deep Learning para simplificar el proceso de reconstrucción, permitiendo un escaneo 3D más accesible y preciso. Con la puesta en marcha del Programa de Referencia y Biobanco Genómico de la Población Argentina PoblAr, que utiliza el protocolo de toma de datos desarrollado e implementado en esta tesis, se espera colectar una gran cantidad de datos de diferentes regiones de Argentina. Se planea correlacionar los datos recolectados con información genética y otros datos asociados, lo que abrirá nuevas posibilidades para comprender la relación entre la genética y la morfología humana, permitiendo tener un impacto significativo en la práctica médica y la investigación biomédica en general.

Técnicas de Inteligencia artificial explicable basadas en una integración de lógica simbólica y no-simbólica

Autor: Pablo Negro[1] https:// orcid .org/ 0000 -0002 -1216 -2651 - pablonegro72@gmail.com

Tutor: Claudia Pons[1, 2, 3] https:// orcid .org/ 0000 -0003 -1149 -0976 - Claudia.Pons@uai.edu.ar

1. CAETI – Centro de altos estudios en tecnología informática, facultad de tecnología. UAI.
2. LIFIA – Facultad de Informática. Universidad Nacional de La Plata.
3. CIC – Comisión de Investigaciones Científicas de Buenos Aires CIC-PBA.

Fecha de Defensa: 16 de julio de 2024

Resumen. La explicabilidad es un aspecto fundamental en el campo del aprendizaje automático, en particular para garantizar la transparencia y la confianza en los procesos de toma de decisiones. A medida que aumenta la complejidad de los modelos de aprendizaje automático, la integración de enfoques neuronales y simbólicos ha surgido como una solución prometedora al problema de la explicabilidad. En este contexto, se ha demostrado que la utilización de métodos de búsqueda para la extracción de reglas en redes neuronales profundas entrenadas es eficaz. Esto implica el examen de los valores de pesos y sesgos generados por la red, generalmente mediante el cálculo de la correlación entre los vectores de pesos y las salidas. La hipótesis desarrollada en este artículo establece que al incorporar la similitud coseno en este proceso, el espacio de búsqueda se puede reducir de manera eficiente para identificar la ruta crítica que conecta las entradas con los resultados finales. Además, para proporcionar una comprensión más completa e interpretable del proceso de toma de decisiones, este trabajo propone la integración de la lógica de primer orden (FOL) en el proceso de extracción de reglas. Al combinar similitud coseno y FOL, se diseñó e implementó un algoritmo innovador que es capaz de extraer y explicar los patrones de reglas aprendidos por una red neuronal entrenada *feedforward*. El algoritmo se probó en tres casos de uso y demostró su eficacia a la hora de proporcionar información sobre el comportamiento del modelo.

Palabras Clave: aprendizaje profundo, entropía, explicabilidad, extracción de reglas, función de distancia, inteligencia artificial, lógica, modelos de caja negra, red neuronal de avance, regularización, similitud de coseno

1 Introducción

Las redes neuronales profundas (DNN) han demostrado una alta precisión predictiva en diversas tareas de aprendizaje automático, siendo las redes *feedforward* entrenadas con *backpropagation* una de las más eficientes. Sin embargo, la representación del conocimiento en estas redes mediante pesos numéricos, sesgos e interconexiones distribuidas dificulta la comprensión por parte del usuario, convirtiéndolas en cajas negras**.** Esto plantea un problema crítico, especialmente en aplicaciones sensibles como el diagnóstico médico**,** donde la falta de explicabilidad puede generar desconfianza o decisiones erróneas**.**

La explicabilidad busca proporcionar una interpretación comprensible del proceso de toma de decisiones del modelo, facilitando la confianza y garantizando su uso seguro y ético en el mundo real. Es un requisito clave en contextos donde la legislación exige explicaciones sobre decisiones algorítmicas que afectan a individuos**.** Diferentes comunidades científicas han abordado este problema desde diversas perspectivas, destacando la extracción de reglas simbólicas como una de las estrategias más relevantes.

En este contexto, este trabajo propone un nuevo método para generar reglas simbólicas que expliquen el comportamiento de una red neuronal artificial**.** La técnica combina similitud coseno con lógica de primer orden**,** basándose en la hipótesis de que los vectores de pesos de la red pueden compararse

mediante esta métrica para identificar qué entradas explican mejor las salidas en cada capa neuronal. Esto permite trazar un camino crítico neuronal**,** representando la función aprendida en términos de reglas lógicas de primer orden**.**
El uso de reglas lógicas permite razonar sobre el comportamiento de la red mediante inferencia lógica, permitiendo determinar salidas en función de entradas específicas. Sin embargo, capturar el comportamiento de las DNN mediante reglas lógicas es un desafío debido a su alta complejidad y al costo computacional del razonamiento lógico en redes grandes.

Este método se enmarca dentro de la **IA Neuro simbólica**, que integra enfoques de IA neuronal y simbólica para superar sus respectivas limitaciones, ofreciendo un enfoque más robusto para el razonamiento, aprendizaje y modelado cognitivo**.**

2 Explicabilidad: Legislación y Aplicación

El uso de métodos basados en datos está impulsando la economía digital y transformando la forma en que investigamos, creamos soluciones y tomamos decisiones. La recopilación y visualización de datos son clave en el descubrimiento científico. En este contexto, la explicabilidad es fundamental para comprender y justificar el comportamiento de las redes neuronales profundas (DNN), asegurando confianza en su uso en el mundo real.
Desde una perspectiva legal, la explicabilidad va más allá del interés científico. Con la adopción del Reglamento General de Protección de Datos (RGPD) en la Unión Europea (https://gdpr-info.eu/), los ciudadanos tienen derecho a una explicación sobre decisiones automatizadas que les afecten. Así, la explicabilidad no solo es un requisito técnico, sino también un derecho jurídico con amplias implicaciones sociales.
Este artículo destaca la importancia de situar al individuo en el análisis de IA aplicada a la toma de decisiones en DNN. Se propone un método basado en lógica de primer orden para extraer reglas comprensibles que expliquen cómo las DNN aprenden y procesan sus entradas. Estas reglas pueden aplicarse directamente a los datos, proporcionando interpretaciones claras del modelo.
Aunque inicialmente diseñado para el mercado de capitales, este método de explicabilidad también beneficia a los equipos de desarrollo de *Deep Learning*, optimizando la selección de hiperparámetros y la topología de las DNN. Además, puede aplicarse en sistemas sanitarios, ayudando a médicos en la toma de decisiones, por ejemplo, en diagnósticos por imágenes mediante la interpretación de las decisiones de la red y las características en las que se basaron.
La solución propuesta presenta una aplicación versátil y escalable, siendo adaptable a diversos problemas que requieren explicabilidad en redes neuronales profundas entrenadas con datos tabulares. Su enfoque permite interpretar el proceso de toma de decisiones de la DNN, proporcionando un marco analítico que puede ser aplicado en múltiples dominios.

3 COLOSSUS: Extracción de reglas con lógica de primer orden para redes neuronales profundas *feedforward* entrenadas.

COLOSSUS significa *Cosmic Logic-Based Rule Extraction for Statistically Understanding Structures* (Extracción de reglas basada en lógica cósmica para comprender estadísticamente las estructuras), lo que refleja el uso de la similitud coseno y métodos estadísticos en este algoritmo.
El algoritmo COLOSSUS es un método de extracción de reglas diseñado con el objetivo de extraer reglas lógicas de DNN *feedforward* entrenadas, combinando lógica de primer orden, similitud coseno y estadísticas obtenidas a partir de los datos de entrada. Las DNN son modelos potentes y complejos que han demostrado un gran éxito en varios campos, pero su falta de interpretabilidad y transparencia

sigue siendo un gran desafío. COLOSSUS aborda este problema al proporcionar un enfoque sistemático para extraer reglas comprensibles e interpretables de las DNN.
La lógica de primer orden es la columna vertebral de este algoritmo, ya que permite traducir las representaciones aprendidas de la DNN en enunciados lógicos. Al asignar los pesos y las activaciones de las neuronas en las capas ocultas a variables y predicados lógicos, el algoritmo crea una representación basada en reglas del proceso de toma de decisiones de la DNN.
La similitud coseno se utiliza para identificar y agrupar neuronas similares, en función de las representaciones aprendidas. Esta agrupación permite que el algoritmo extraiga reglas más generales, en lugar de reglas individualizadas para cada neurona. Al considerar la similitud entre neuronas, el algoritmo puede capturar las características importantes de los datos de entrada que contribuyen a la decisión final de la red neuronal.
Por último, el algoritmo utiliza estadísticas para refinar las reglas extraídas. Evalúa la frecuencia y la importancia de cada regla para determinar su relevancia e importancia en el proceso de toma de decisiones de la red neuronal. Este paso garantiza que solo se extraigan las reglas más relevantes y significativas, lo que da como resultado un conjunto de reglas más conciso e interpretable.
En general, COLOSSUS combina las potentes capacidades de representación de la lógica de primer orden, la capacidad de capturar patrones generales mediante similitud coseno y el refinamiento de reglas mediante estadísticas para extraer reglas interpretables de las DNN. Esto permite una mejor comprensión del proceso de toma de decisiones de las redes neuronales y puede ayudar a mejorar su rendimiento y confiabilidad.

3.1 Análisis de Contexto

La tarea de extracción de reglas puede verse como una tarea de búsqueda o como una tarea de aprendizaje. En el enfoque de búsqueda, las reglas se extraen a nivel de neuronas individuales (ocultas y de salida) en la red, observando sus pesos y sesgos. Así, uno de los principales problemas de este enfoque es cómo restringir el espacio de búsqueda para las posibles combinaciones de pesos y sesgos. En una red neuronal entrenada, el conocimiento adquirido en la fase de entrenamiento se codifica en la arquitectura de la red, las funciones de activación utilizadas y los pesos y sesgos de las neuronas. En este sentido, el rendimiento de una red neuronal está directamente relacionado con su arquitectura y parámetros. Por lo tanto, la elección de una arquitectura influye en el tiempo de aprendizaje, la precisión predictiva, la tolerancia al ruido y la capacidad de generalización de la red.
En el presente trabajo, la tarea de extracción de reglas consiste en analizar los vectores de pesos generados en cada neurona, junto con el vector de sesgo de cada capa de la red, y extraer un conjunto de reglas explicables mediante lógica de primer orden. Consideramos casos en los que las entradas de la red neuronal *feedforward* son tabulares y las salidas son categóricas (es decir, problemas de clasificación).
El método de propagación de neuronas en redes neuronales *feedforward* se define por la siguiente forma canónica:

$$Z^{[l]} = W^{[l]}A^{[l-1]} + b^{[l]} \qquad \text{eq.1}$$

$$A^{[l]} = g^{[l]}(Z^{[l]}) \qquad \text{eq.2}$$

Donde:

$W^{[l]}$ = representa la matriz de pesos para las neuronas de la capa I.
$A^{[l-1]}$ = representa los valores de entrada de la capa l, donde X = $A^{[0]}$.
$b^{[l]}$ = es el vector de *bias* para la capa l.
$g^{[l]}$ = es la función de activación de la capa l.
$A^{[l]}$ = contiene los valores de salida de las neuronas de la capa I.

La función de activación para capas ocultas tiene activaciones definidas por:

$$ReLU = A^{[l]} = \max(0, x) \qquad \text{eq.3}$$

Y para la capa de salida definida por:

$$SoftMax \text{ function} = f(z)_j = \frac{e^{zj}}{\sum_{n=1}^{m} e^{zn}} \qquad \text{eq.4.1}$$

O

$$Sigmoid \text{ function} = f(z) = \frac{1}{1+ e^{-z}} \forall\, z\, real \qquad \text{eq.4.2}$$

3.2 Alcance

Este trabajo se centra en DNN clásicas *feedforward* entrenadas, ya que el enfoque está en la extracción de reglas a partir de sus matrices de pesos. Variaciones de este tipo de redes, como las redes neuronales recurrentes y los Transformers, que, aunque hacen uso de redes neuronales *feedforward*, presentan arquitecturas diferentes al utilizar capas de atención, *embeddings* y transformaciones, creando arquitecturas híbridas más complejas, quedan fuera de este estudio.
Por lo tanto, al referirnos a tecnologías de *Deep Learning*, hacemos referencia a DNN que se describen como capas de neuronas densas, combinadas con capas de entrada y salida, y alguna función de activación en sus neuronas.

3.3 El método propuesto

Esta sección explica cómo funciona la combinación de la similitud del coseno y las técnicas de lógica de primer orden para extraer reglas de una red neuronal profunda *feedforward* entrenada. Los pesos y sesgos se tratan con la misma precisión y signos con los que son generados por la red, es decir, no se les aplica ninguna transformación. A modo ilustrativo en el presente artículo, y para ejemplificar el funcionamiento del algoritmo de extracción de reglas, se presenta un solo caso de uso de DNN *feedforward* de tres capas entrenadas mediante *backpropagation*.
Se considera que una neurona en cada capa es crítica cuando el cálculo de la métrica de similitud del coseno con respecto a los datos de entrada y los pesos de la neurona supera un determinado valor umbral. Se explicará en detalle el procedimiento para extraer una regla de confirmación.
Como se mencionó en la Sección 1, uno de los problemas más cruciales en el desarrollo de un algoritmo de extracción de reglas es cómo restringir el tamaño del espacio de soluciones buscado. Cada neurona dentro de la red entrena un número N de pesos, donde N es el número de entradas de cada neurona. Se puede pensar en N como el número de neuronas en la capa $L^{[l-1]}$. Así, en una pasada *feedforward*, para la capa $L^{[l]}$ tendremos N entradas en cada neurona y M salidas, donde M es el número de neuronas en la capa $L^{[l]}$. Cada valor en las salidas de las neuronas de la capa $L^{[l-1]}$ se presenta a cada neurona de la capa $L^{[l]}$, donde se procesan mediante multiplicaciones y sumas y luego se exponen a alguna función de activación. Esto se expresa en la ecuación 1 y 2.
Luego, se calculará la similitud coseno entre el vector $Z^{[l]}$ de la ecuación 1 y el vector de pesos formado por el peso que cada neurona asignó a la característica F_N. Llamaremos a este vector $T^{[l]}{}_N$.
Esto nos permitirá entender qué características de entrada explican mejor los resultados. De esta manera, si después de aplicar el cálculo de similitud coseno entre ambos vectores, la métrica supera el valor umbral, diremos que la neurona asociada con la característica F_N es crítica, y se incluirá en el grupo de neuronas de la capa $L^{[l]}$ que explican la función que la red aprendió para una clase dada.
La siguiente sección explica en detalle cómo la comparación de vectores ayuda a determinar la ruta crítica y a extraer la función aprendida por la red neuronal profunda.

3.4 Similitud Coseno

La similitud coseno es una medida de la similitud entre dos vectores en un espacio que posee un producto interno, con el cual se evalúa el valor del coseno del ángulo entre ellos. Esta función trigonométrica proporciona un valor igual a 1 si el ángulo incluido es cero, es decir, si ambos vectores apuntan en la misma dirección. Para cualquier otro ángulo existente entre los vectores, el coseno dará un valor menor que uno. Si los vectores fueran ortogonales, el coseno se anularía, y si apuntaran en direcciones opuestas, su valor sería -1. De esta manera, el valor de esta métrica se encuentra entre -1 y 1, es decir, en el intervalo cerrado [-1,1]. Dado dos vectores A y B de dimensión N, la similitud del coseno se calcula como:

$$similarity = \cos(\emptyset) = \frac{A.B}{||A||.||B||} = \frac{\sqrt{\sum_{i-1}^{n} A_i B_i}}{\sqrt{\sum_{i-1}^{n} A_i^2}\sqrt{\sum_{i-1}^{n} B_i^2}} \qquad \text{eq.5}$$

3.5 Número Mínimo de Neuronas

Como se explica en la siguiente sección, durante la segunda etapa del algoritmo COLOSSUS, se identifican las neuronas críticas en toda la red para determinar un camino neuronal crítico por clase (en el caso de tareas de clasificación) y, de este modo, ser capaces de extraer las funciones que la red ha aprendido y explicarlas mediante lógica proposicional.
Para la capa cero, donde las neuronas de esta capa están mapeadas a los datos de entrada, existe un conjunto mínimo de neuronas necesarias (MNN) para que el algoritmo que explica las reglas alcance una alta precisión (acc≥90). Este conjunto MNN se puede calcular de dos maneras:

1.	(NFDS/2) + 1	Donde **NFDS** es el número total de características en el conjunto de datos. Para el caso en el que el número de neuronas en la capa de entrada, sea menor que 10.
2.	NFDS / QOL	Donde QOL es el número de neuronas en la capa de salida.

En el caso en que el método de extracción de reglas identifique menos neuronas que el conjunto MNN, es posible seleccionar neuronas adicionales de la siguiente manera:

- Calcular la entropía para las características que se asignan a las neuronas del conjunto secundario identificado en la etapa 2 del algoritmo.
- Si el conjunto secundario de neuronas es un conjunto vacío, entonces la entropía se calcula sobre todas las características no críticas.
- Seleccionar el número de neuronas faltantes tomando como criterio de selección las características que tienen la entropía más alta.
- Las neuronas adicionales se agregan al conjunto de neuronas críticas por clase para $A^{[0]}$ y se tienen en cuenta en el cálculo de la disyunción explicado en la tercera etapa del algoritmo en la siguiente sección.

Dado que cada característica se asigna a una neurona en la capa de entrada, estas neuronas se marcarán como críticas y se tendrán en cuenta en el algoritmo de validación de reglas. Tenga en cuenta que esta adición de neuronas es opcional.

3.6 El algoritmo COLOSSUS para la extracción de reglas

El algoritmo de extracción de reglas COLOSSUS se basa en la estrategia de comparación vectorial de pesos discutida anteriormente y funciona en cuatro etapas.

En la primera etapa, se calculan estadísticas para los datos de entrada para las cuales la precisión predictiva de la red se maximiza. Dado que COLOSSUS busca aquellos valores que maximicen la precisión de la red, se buscaran las neuronas criticas dentro de la red con dichos valores estadísticos. Además, las estadísticas ayudaran en la verificación de las reglas que expliquen a los datos. Estas determinaran las cotas para cada *feature* critico de entrada.
En la segunda etapa se calcula la similitud coseno entre el vector $Z^{[l]}$ de la eq.1 con el vector $T^{[l]}$ de pesos formado por el peso que cada neurona le asigno al *feature* $R^{L}_{(F)}$ con el objetivo de entender cuales *features* de entrada mejor explican los resultados. De este modo, si después de aplicar el cálculo de similitud coseno entre ambos vectores la métrica supera el valor de umbral, diremos que la neurona asociada al *feature* $R^{L}_{(F)}$ es crítica, y se la incluirá en el grupo de neuronas de la capa R^{L} que permitirán explicar la función que aprendió la red para una clase dada. Este proceso se repite para cada neurona de cada capa.
Utilizando la misma técnica, se calcula la similitud coseno entre los vectores de pesos de las diferentes neuronas en cada capa, con el objetivo de entender cuales neuronas guardan una relación entre sus sistemas de pesos. Entonces si una neurona critica guarda una relación vectorial con una o más neuronas, se dirá que estás neuronas trabajan en conjunto, y se hará referencia a estas como el conjunto de neuronas secundarias.
Se analiza el factor MNN, según lo expuesto en la sección 3.5, para completar el conjunto de neuronas críticas.
En la tercera etapa, se calculan las relaciones de conjunción entre las neuronas críticas de cada capa para todas las clases, con el objetivo de obtener una única tabla de verdad por capa.
Para calcular las relaciones de conjunción se utilizan los vectores de peso de las neuronas críticas. Si el valor del peso es positivo el valor será V o 1, y si es negativo, será F o 0. Así, si, por ejemplo, para una determinada capa, se identifican dos neuronas como críticas, la conjunción será calcularse como $R^{[L]}_{n1} \wedge R^{[L]}_{n2} = TT^{[L]}$.
Una vez obtenidas las tablas de verdad por capa, se calcula la relación de implicancia entre las tablas de verdad de cada capa ($TT^{[L-1]} \rightarrow TT^{[L]} \rightarrow TT^{[L+1]} = TTdnn$). El objetivo de calcular las implicancias es comprobar que los resultados sean tautologías, y asegurarnos de que las neuronas marcadas como críticas son válidas y así garantizar que el *camino neuronal critico* encontrado tiene un fundamento basado en lógica valido. Las reglas obtenidas se escriben en términos de fórmulas bien formadas (fbf) de la lógica proposicional para obtener reglas que expliquen la relación entrada-salida. Con estas fbf se explican las funciones que aprendió la red para cada clase.
Es importante notar que, además de las fbf extraídas, también podemos identificar una subred regularizada que explica la clase que se esté analizando.
En la etapa final, las neuronas identificadas como críticas en la capa de entrada $A^{[0]}$ se mapean con los datos estadísticos obtenidos en la primera etapa, para determinar la precisión de estas de las reglas, y se calcula una función de distancia para resolver posibles solapamientos entre clases. Como resultado de esta etapa se obtienen fórmulas de lógica de primer orden, que explican los datos de entrada. El algoritmo de validación se explica en detalle en la siguiente sección.

4 Algoritmo de validación de reglas

Como se mencionó anteriormente, en la primera etapa del algoritmo de extracción de reglas, se calculan e identifican las estadísticas que maximizan la precisión de la red para cada clase. Por lo tanto,

el algoritmo de validación de reglas utiliza dichas estadísticas por clase y por característica para emplearlas como valores de dimensión.
Es importante destacar que el algoritmo de validación de reglas opera sobre la **unión** de las características identificadas como críticas para la capa $A^{[0]}$ en cada clase. Es decir, si un conjunto de datos tiene **3 clases posibles**, y el algoritmo de extracción identifica para la capa $A^{[0]}$ las siguientes neuronas por clase:

$$\text{Class } 1 = N_0 \qquad \text{Class } 2 = N_0 \text{ ^ } N_2 \qquad \text{Class } 3 = N_1$$

El algoritmo de validación operará sobre la unión de las características de todas las clases: if($N0$ ^ $N1$ ^ $N2$) ➔ClassN.

Como siguiente paso, se lee e itera sobre todo el conjunto de datos, verificando que cada característica relevante se encuentre dentro del rango de valores delimitado por las estadísticas obtenidas en la primera fase. Si la verificación es positiva, se marca dicha instancia (registro dentro del conjunto de datos) como perteneciente a una determinada clase.
Si los datos en el conjunto de datos presentan un alto solapamiento entre sí, puede ocurrir que el algoritmo, para la misma instancia de los datos de entrada, indique que dicha instancia pertenece a más de una clase. En este caso, la **clase correcta** se determinará mediante el cálculo de una función de distancia que evalúa qué tan alejada está la característica con **mayor entropía** con respecto al **centroide** de la característica que las clases tienen en común para las clases superpuestas, de la siguiente manera:

- **fme** = Es la variable que tiene la mayor entropía dentro del conjunto de características seleccionadas.
- **fcc** = Es una característica que ambas clases tienen en común (dentro del conjunto de características seleccionadas). Si no hay características en común o si hay más de una, se elige la que tenga la mayor entropía.

Para este ejemplo, se supone que las clases superpuestas son las clases 0 y 1.

- dst0 = (fme – MinVal_Class0(fcc)) + (MaxVal_Class0(fcc) - fme) eq. 6
- dst1 = (fme – MinVal_Class1(fcc)) + (MaxVal_Class1(fcc) - fme) eq. 7

Class = (1 if dst1 < dst0 else 0) eq. 8 ➔ La desigualdad puede variar dependiendo de la función de activación utilizada en la capa de salida.

Es importante destacar que para identificar las neuronas y caminos neuronales críticos para cada clase se utilizan estadísticas que maximizan la precisión de la red. Para el algoritmo de validación de reglas y para un mayor grado de generalización, las estadísticas para cada característica se toman de sus valores máximos y mínimos.
Finalmente, para medir la precisión del algoritmo de extracción de reglas, se compara la clase dada *Class* con la verdad fundamental proporcionada en el conjunto de datos.

5 Caso de Estudio

En esta sección se presentan los resultados de la prueba del algoritmo COLOSSUS en el conjunto de datos Wine.

El conjunto de datos Wine incluye información sobre tres clases de vinos producidos en Italia y 13 características predictivas. Este conjunto de datos es el resultado de un análisis químico de vinos cultivados en la misma región de Italia, pero provenientes de tres cultivares diferentes. El análisis determinó las cantidades de 13 componentes presentes en cada uno de los tres tipos de vinos.
El conjunto de datos contiene 177 ejemplos, distribuidos de la siguiente manera: 59 ejemplos de vino tinto, 71 ejemplos de vino rosado, 48 ejemplos de vino blanco.
Los *features* de entrada a la red para este conjunto de datos son los siguientes: *Alcohol, Malic acid, Ash, Alkalinity of Ash, Magnesium, Total Phenols, Flavonoids, Nonflavanoids Phenols, Proanthocyanins, Color Intensity, Hue, Dilution, Proline.*
La salida esperada es obtener la mayor precisión posible en la predicción de la clase del vino usando datos de prueba, diferenciando entre tinto, rosado y blanco.
Para este caso de uso, se configuró una DNN con la siguiente estructura: Capa de entrada: 13 neuronas (una por cada característica). Capas ocultas: 2 capas con 12 neuronas cada una. Función de activación para capas ocultas: *ReLU*. Capa de salida: 3 neuronas (una para cada clase de vino). Función de activación de la capa de salida: *SoftMax*. Umbral para determinar si una neurona es crítica: 0.6.
La red neuronal entrenada con este conjunto de datos alcanzó una precisión del 92%.
Tras aplicar los pasos identificados en las etapas 2 y 3 del algoritmo, se obtuvieron las siguientes fórmulas bien formadas (fbf) para cada clase, antes de aplicar el factor MNN.

Class 0 = $N^0_{(12)}$ -> $N^1_{(10)}$ -> $N^2_{(2)} \wedge N^2_{(5)} \wedge N^2_{(8)} \wedge N^2_{(9)}$
Class 1 = $N^0_{(12)}$ -> $N^1_{(1)} \wedge N^1_{(10)}$ -> $N^2_{(5)} \wedge N^2_{(8)} \wedge N^2_{(9)}$
Class 2 = $N^0_{(12)}$ -> $N^1_{(10)}$ -> $N^2_{(5)} \wedge N^2_{(8)} \wedge N^2_{(9)}$

La característica que el algoritmo identificó como crítica es *proline* (X_{12}). Como se explicó en el punto 3.5, es posible aumentar el número de características consideradas críticas con el objetivo de mejorar la precisión en la explicación de los datos.
Para un MNN = 2, si dividimos el número de características (13) por el número de neuronas en la capa de salida (3), obtenemos un MNN = 4. Esto significa que podemos agregar 3 características adicionales al conjunto de características críticas.
Dado que, en este caso de uso, el conjunto de neuronas secundarias es un conjunto vacío, la entropía se calcula sobre todas las características no críticas.
El análisis identifica las siguientes características como críticas y susceptibles de ser agregadas al conjunto de neuronas críticas: *flavanoids* (X_6), *color intesity*(X_9), *malic acid* (X_1)

Para el cálculo de la distancia se utilizaron las características *prolina* (X_{12}) y *flavonoides* (X_6) ya que la primera es la característica que fue identificada por el algoritmo de extracción y es común a todas las clases, y la segunda es la que presenta mayor entropía, por lo tanto, las métricas para el cálculo de la distancia se expresan de la siguiente manera:

dst_X = ($proline_{(N)}$ – MinValue_ClassX(*'flavanoids'*)) + (MaxValue_ClassX(*'flavanoids'*) - $proline_{(N)}$)
por ➔ eq. 6

dst_Y = ($proline_{(N)}$ – MinValue_ClassY(*'flavanoids'*)) + (MaxValue_ClassY(*'flavanoids'*) - $proline_{(N)}$)
por ➔ eq. 7

Después de aplicar el cálculo de la eq. 8, obtenemos Class = (X if dstX < dstY else Y).

Finalmente, el valor obtenido en *Class* se compara con el *ground truth* para la instancia N. Entonces las reglas obtenidas para las diferentes clases del conjunto de datos Wine serían las siguientes:

Class0 = ($\forall X_1$[1,35..4,04] $\wedge$ $\forall X_6$[2,19..3,93] $\wedge$ $\forall X_9$ [3,52..8,9] $\wedge$ $\forall X_{12}$[680..1680])
$\wedge$ fnc_dst_class($proline_{(N)}$, *flavanoids*, 0, N) == 0

Class1 = ($\forall X_1$[0,74..5,8] $\wedge$ $\forall X_6$[0,57..5,08] $\wedge$ $\forall X_9$ [1,26..6] $\wedge$ $\forall X_{12}$[278..985])
$\wedge$ fnc_dst_class($proline_{(N)}$, *flavanoids*, 1, N) == 1

Class2 = ($\forall X_1$[1,24..5,65] $\wedge$ $\forall X_6$[0,34..1,57] $\wedge$ $\forall X_9$ [3,85..13] $\wedge$ $\forall X_{12}$[415..880])
$\wedge$ fnc_dst_class($proline_{(N)}$, *flavanoids*, 2, N) == 2

fnc_dst_class(fme, fcc, ClassX, ClassY) = $\exists \boldsymbol{X_n}$ ($dst_x < dst_y$)

Para este conjunto de datos, el algoritmo de explicación de reglas logra una precisión del 97%.

6 Lineas de trabajo futuro

Las futuras líneas de investigación se resumen en los siguientes puntos:

- Investigar la aplicación del método propuesto en redes neuronales más complejas y a mayor escala, con el objetivo de evaluar su escalabilidad y rendimiento.
- Realizar estudios en conjuntos de datos industriales más grandes, que requieran el desarrollo de arquitecturas de redes neuronales profundas más sofisticadas.
- Explorar la integración de marcos lógicos adicionales, como lógica difusa o lógica probabilística, para mejorar la interpretabilidad y explicabilidad de los patrones de reglas extraídas.
- Llevar a cabo estudios empíricos para evaluar la efectividad del método en diferentes tareas y dominios, como mercados de capitales, banca y salud, con el fin de analizar su generalización, precisión y robustez. Además, realizar comparaciones con otros modelos para evaluar la precisión, legibilidad e interpretabilidad de las reglas extraídas.
- Examinar el impacto de diferentes arquitecturas de red y técnicas de entrenamiento en la interpretabilidad de los patrones de reglas extraídas, para comprender mejor la relación entre la estructura de la red y su capacidad de explicación.
- Investigar el potencial del método propuesto en combinación con otras técnicas de explicabilidad, como mecanismos de atención o mapas de prominencia (*saliency maps*), para proporcionar una comprensión más completa y detallada del proceso de toma de decisiones en redes neuronales.

7 Contribuciones Principales

Las principales contribuciones de este trabajo son:

- Se propone un nuevo algoritmo para extraer los patrones de reglas aprendidos por una red neuronal *feedforward* entrenada y analizar sus propiedades.
- Se emplean tanto la lógica proposicional como la lógica de primer orden (FOL) para explicar los patrones aprendidos por la red neuronal.
- Se presenta un método de caja blanca que analiza las matrices de pesos de redes neuronales profundas (DNNs) para extraer reglas, utilizando la métrica de similitud coseno.
- Se implementa una función de distancia para resolver solapamientos entre clases.
- El método no requiere reglas de aprendizaje especiales durante la extracción de reglas, lo que facilita su uso.
- El método es eficiente tanto para atributos continuos como enumerados.

- El algoritmo identifica de manera eficiente las neuronas relevantes de la DNN y no requiere reentrenamiento tras la selección de neuronas críticas.
- Las reglas extraídas permiten explicar el proceso de toma de decisiones de la red neuronal mediante fórmulas de lógica de primer orden, lo que puede ser útil en diversas aplicaciones como diagnóstico médico, detección de fraudes y evaluación crediticia.
- Las reglas extraídas se mapean directamente al conjunto de datos, lo que permite explicar las decisiones de la red en función de los datos.
- De manera indirecta, el algoritmo de extracción de reglas propone un método de regularización para la DNN, identificando los caminos neuronales críticos, lo que da lugar a redes más pequeñas y optimizadas.

8 Conclusiones

Este artículo describe el algoritmo COLOSSUS, el cual genera reglas simbólicas para explicar el comportamiento de una red neuronal artificial. El algoritmo comienza con un proceso de cálculo de estadísticas sobre los datos y, posteriormente, extracción de pesos de la red neuronal entrenada, donde se calcula la similitud coseno entre los vectores. Este cálculo permite establecer la relación de cada variable de entrada con cada neurona, entre sí, con el vector de sesgo y con el vector formado por los valores de activación en cada capa. Además, se aplica el mismo procedimiento a los vectores de pesos de cada neurona, lo que da lugar a una comparación interneuronal de los vectores de pesos dentro de la misma capa, con el objetivo de comprender la colaboración entre las neuronas de una determinada capa.

Las reglas resultantes se presentan como fórmulas bien formadas de lógica proposicional, proporcionando una visión clara de la correlación entre las variables de entrada y la salida de la red. Este método identifica caminos neuronales críticos mediante el análisis del flujo de datos desde la entrada hasta la salida y genera redes regularizadas para cada clase.

Las reglas extraídas se basan en las estadísticas obtenidas de los datos de entrada y se seleccionan por su alta precisión al explicar los resultados de la red. La alineación entre estas métricas sirve como base para validar y verificar las reglas extraídas. Las neuronas identificadas en la capa de entrada por el algoritmo se mapean a las características del conjunto de datos, donde el algoritmo de validación de reglas opera sobre estas características y las estadísticas previamente recopiladas. Además, se emplea una función de distancia para resolver solapamientos entre clases.

Una de las principales fortalezas de esta técnica es su simplicidad y eficiencia, ya que no requiere reglas de aprendizaje previas para la extracción de reglas.

También es adaptable a diferentes tipos de atributos y produce reglas completas y válidas, además de identificar neuronas importantes y caminos neuronales críticos.

Gracias al uso de lógica de primer orden, las reglas extraídas pueden interpretarse fácilmente, lo que las hace altamente valiosas para diversas aplicaciones.