



POSTGRADO | INFO

# TESIS Y TESISISTAS

2023

DOCTORADO  
MAESTRÍAS  
ESPECIALIZACIONES



Facultad de  
**INFORMÁTICA**



UNIVERSIDAD  
NACIONAL  
DE LA PLATA

**INDICE**

**TESIS Y TESISISTAS**  
2023

<b>EQUIPO EDITORIAL</b>	<b>P. 4</b>
<b>NOTA EDITORIAL</b>	<b>P. 5</b>

## DOCTORADO EN CIENCIAS INFORMÁTICAS

<b>Dr. Juan Carlos Sandi Delgado</b>	<b>P. 8</b>
Propuesta Metodológica y Arquitectónica para el Desarrollo de Juegos Serios Orientados a la Formación de Competencias Tecnológicas	
Methodological and architectural proposal to develop serious games oriented to the formation of technological skills	
<b>Dr. Juan Manuel Rodríguez</b>	<b>P. 12</b>
Integración de métodos de descubrimiento de conocimiento embebido en fuentes de información desestructuradas	
Integration of Embedded Knowledge Discovery Methods in Unstructured Information Sources	
<b>Dr. Juan Cruz Gardey</b>	<b>P. 14</b>
Métodos y herramientas de desarrollo y evaluación de refactorings para la mejora de la experiencia de usuario en aplicaciones web	
Methods and tools for refactoring development and evaluation to improve users experience in web applications	
<b>Dr. Jorge Iván Pincay Ponce</b>	<b>P. 18</b>
Análisis de datos educativos aplicado en el estudio de la incidencia de factores socioeconómicos en el rendimiento escolar	
Analysis of educational data applied in the study of the incidence of socioeconomic factors in school performance	
<b>Dr. Luis Orlando Philco Asqui</b>	<b>P. 22</b>
Calidad de sensado y eficiencia energética en redes de sensores inalámbricos (WSN)	
Quality of sensing and energy efficiency in Wireless Sensor Networks (WSN)	
<b>Dr. Daisy Elizabeth Imbaquingo Esparza</b>	<b>P. 24</b>
Método de auditoría informática para minimizar el riesgo de calidad de los resultados basado en sistemas de procesamiento avanzado de datos	
Computer Audit Method Based on Advanced Data Processing Systems to Minimize the Risk of quality issues	
<b>Dr. Juan Pablo Tessore</b>	<b>P. 26</b>
Modelado e implementación de algoritmos inteligentes de análisis de opinión	
Modeling and implementation of opinion analysis smart algorithms	

## Dr. Guido Sebastián Tebes

P. 30

Estrategia Integrada de Pruebas de Software consciente de la Situación y basada en Escenarios

Situation-aware Scenario-based Integrated Software Testing Strategy

## Dr. Martín Becerra

P. 34

Framework para la explotación de procedimientos semánticos para navegadores de RA

Framework for Interoperable Dynamic Procedure interactions on Semantic Augmented Reality Browsers

## MAESTRÍAS

### INTELIGENCIA DE DATOS ORIENTADA A BIG DATA

#### Mg. Fernández Juan Manuel

P. 38

Clasificación automática de correos electrónicos

Automatic email classification

### CÓMPUTO EN ALTAS PRESTACIONES

#### Mg. Leandro Ariel Libutti

P. 42

Coplanificación de procesos maleables de aprendizaje automático mediante contenedores

Scheduling elastic machine learning process through containers

### INGENIERÍA DE SOFTWARE

#### Mg. Cecilia Alicia Moreyra

P. 46

Soporte al proceso de análisis e interpretación de variantes genéticas en Exoma humano

Support to the process of analysis and interpretation of genetic variants in human Exome

#### Mg. Adolfo Tomás Spinelli

P. 48

Elicitación de requerimientos centrada en el usuario para el desarrollo de Serious Games

User-based requirements elicitation to develop serious games

#### Mg. Víctor Daniel Martínez

P. 50

Minería de Procesos en la Ingeniería de Software Análisis del proceso de distribución de piezas postales

Process mining in Software Engineering. An analysis of the postal distribution process

### REDES DE DATOS

#### Mg. Marcelo Alberto Gómez

P. 52

Seguridad en el enrutamiento utilizando tecnología Blockchain

Routing security using Blockchain technology

## TECNOLOGÍA INFORMÁTICA APLICADA EN EDUCACIÓN

#### Mg. Leonardo José Navarría

P. 54

Laboratorio virtual de electrónica básica para alumnos universitarios dentro de aula extendida

Verifications of the basic laws of electric circuits and applications on semiconductors

#### Mg. María Paula Dieser

P. 58

Autorregulación del Aprendizaje, Estrategias de Participación y Rendimiento Académico. Un estudio de su relación en comunidades virtuales de indagación

Learning self-regulation, participation strategies and academic performance. A study on how they relate in virtual communities of inquiry.

## ESPECIALIZACIONES

### INTELIGENCIA DE DATOS ORIENTADA A BIG DATA

#### Esp. Alejo Hernández

P. 64

Bases de datos de series temporales y métodos estadísticos para la predicción de rendimientos de ETH

Database of time-series and statistical methods to predict ETH performance

#### Esp. Hugo Manuel Fajardo

P. 68

Estudio Comparativo entre Apache Spark y Apache Flink en el procesamiento de streaming en entornos Big Data

A comparison between Apache Spark and Apache Flink and how they process streaming in Big Data environments

#### Esp. Mercedes Barrionuevo

P. 70

Análisis Visual de Datos Temporales Asociados a Criptomonedas

Visual Analysis of Temporary Data Associated with Cryptocurrencies

### INGENIERÍA DE SOFTWARE

#### Esp. Luciano Marrero

P. 72

Estudio Comparativo de BD NO SQL

BD, NO and SQL: a comparison

## TECNOLOGÍA INFORMÁTICA APLICADA EN EDUCACIÓN

#### Esp. Leandro Javier Castro

P. 74

Programación tangible en el ámbito educativo. Análisis de experiencias

Tangible programming in education. Analyzing experiences

#### JURADOS DESIGNADOS 2023

P. 77

# EQUIPO EDITORIAL

## **DIRECTOR DE POSTGRADO**

Dr. R. Marcelo Naiouf

## **SECRETARIA DE POSTGRADO**

Dra. Laura De Giusti

## **DIRECTORA ADMINISTRATIVA DE POSTGRADO**

Lic. Alejandra Pizarro

## **SECRETARÍA ADMINISTRATIVA DE POSTGRADO**

Natalia Otero

Débora Mieres

Maitén Meza

Carolina Covas

Valentín Altavista

Abril Buffarini

Florencia Rugera

Soledad Bravo

Emilia Zubieta

## **COORDINADOR DE POSTGRADO**

Ing. Armando De Giusti

# NOTA EDITORIAL

En un trabajo conjunto de la Secretaría de Postgrado, la Dirección de Postgrado y la Secretaría Administrativa de Postgrado se presenta el Libro de Tesis y Tesistas correspondiente al año 2023, en la Facultad de Informática de la UNLP.

En sus páginas se pueden recorrer la síntesis de las Tesis de Doctorado y Maestría, así como los Trabajos Finales de Especialización aprobados a lo largo del año 2023.

Acompañamos estos resúmenes de los datos de los autores, de sus directores y también un listado de los distinguidos profesores e investigadores de Argentina y de Universidades del exterior que han sido Jurados en las Tesis.

En los 22 trabajos que se sintetizan y cuyos contenidos completos se pueden obtener del repositorio institucional de la UNLP (SE-DICI) se refleja el intenso trabajo que desarrolla el área de Postgrado de la Facultad de Informática, trabajo iniciado en 1995 en la Facultad de Ciencias Exactas y consolidado a partir de 1999 en la Facultad de Informática.

Es interesante resaltar que de estas 22 Tesis, 3 corresponden a docentes-investigadores con cargos en la Facultad de Informática de la UNLP y 19 no tienen relación directa con nuestra Facultad.

Con este libro esperamos contribuir a la difusión de las actividades académicas de Postgrado, así como a la trasmisión de conocimientos generados en la Facultad de Informática de la UNLP.

## **Equipo Editorial del Postgrado**



01

DOCTORADO EN  
**CIENCIAS INFORMÁTICAS**



Directora  
Dra. Patricia Alejandra Bazán

Fecha de aprobación  
15 de junio de 2023

SEDICI  
[Tesis Completa](#)

## **Propuesta Metodológica y Arquitectónica para el Desarrollo de Juegos Serios Orientados a la Formación de Competencias Tecnológicas**

**Palabras clave:** Juegos serios; formación de competencias tecnológicas; prototipo; profesorado; propuesta metodológica y arquitectónica; INTEGRA+506; SkillNet; redes de datos.

### **Motivación**

Ante el surgimiento de las de las Tecnologías de la Información y de la Comunicación (TIC) y surápido avance e impacto a nivel educativo, las Instituciones de Educación Superior (IES) se ven obligadas a enfrentarse al reto de mantenerse actualizadas ante las nuevas demandas de la sociedad actual (Vega-Hernández et al., 2018). Asimismo, (Cabero et al., 2015) señalan que las TIC brindan alas IES el potencial para la creación de escenarios de aprendizaje que permiten el acceso inmediato tanto de información como de recursos tecnológicos que facilitan la formación permanente de estudiante y el profesorado en tecnologías digitales (Es importante aclarar que por razones culturales se utiliza en Costa Rica la palabra “docente” como sinónimo de “profesor” o “profesorado”. Por ende, en este documento se utilizan ambos términos como sinónimos).

Aguaded & Cabero (2014), afirman que el principal problema que ha manifestado el profesorado es la actualización profesional, ya que requieren de diversas actividades de capacitación relacionadas a potenciar sus habilidades y competencias tecnológicas, mismas requeridas para utilizar e integrar las TIC en los procesos formativos.

En investigaciones realizadas por Teo (2009) y Scherer et al., (2015), se afirma que el proceso de enseñanza y aprendizaje apoyados por las TIC pueden ser innovados mediante la incorporación de juegos serios en las metodologías de trabajo. Lo anterior, debido a que los mismos pueden ser utilizados para la adquisición de habilidades, formación de competencias, cambios de actitud, comportamiento y, pueden potenciar la motivación y el rendimiento del aprendizaje, en esta línea, los juegos serios pueden adquirir un rol protagónico ante la posibilidad de convertirse en un medio para el acercamiento del profesorado a las tecnologías digitales (Boyle et al., 2011; Chang et al., 2012).

Sandí & Bazán (2017) realizaron una investigación con el objetivo de plantear retos y desafíos para la ejecución de juegos serios educativos utilizando las posibilidades que brinda el modelo de SaaS en Cloud Computing. Como resultado del trabajo se obtuvo una lista de características funcionales que los juegos serios educativos adquirirían al ser utilizados como software como servicio.

En esta misma línea, Sandí & Sanz (2018) llevaron a cabo una investigación con el propósito de indagar el estado del arte de los juegos serios educativos y sus posibilidades para la adquisición de competencias tecnológicas en la formación del profesorado. Como resultado se obtuvo la elaboración de una propuesta de integración de juegos serios para la formación de competencias tecnológicas en el profesorado y para la mejora de sus actitudes frente a estas tecnologías. Además, se elaboró de una propuesta para la evaluación del grado de aceptación de las tecnologías digitales por parte del profesorado, en particular sobre juegos serios.

Si bien existe variedad de investigaciones que hacen referencia a la posibilidad que ofrecen los juegos serios para la formación de habilidades y competencias digitales, no se evidencian juegos serios diseñados exclusivamente para potenciar habilidades y competencias tecnológicas del profesorado a nivel de educación superior. Asimismo, no se ha localizado o identificado en las referencias bibliográficas analizadas, la existencia de metodologías específicas que orienten o sirvan de guía para la creación y diseño de juegos serios orientados a la formación de competencias digitales.

Los resultados obtenidos en las investigaciones realizadas dan pie y son referente teórico para la elaboración de una investigación que permita proponer una metodología y arquitectura de software que guíe los pasos a seguir durante la creación e implementación de juegos serios que vayan a ser



utilizados para la formación de habilidades o competencias tecnológicas. En consecuencia, se plantea aplicar la metodología y la arquitectura de software que se proponga para crear y diseñar un prototipo no funcional de juego serio que pueda utilizarse para la formación de habilidades o competencias tecnológicas.

a) Objetivo general

- Definir una metodología y arquitectura de software para el diseño de juegos serios que permitan la adquisición o formación de competencias tecnológicas.

b) Objetivos específicos

- Estudiar los conceptos básicos vinculados con las TIC, competencias tecnológicas, juegos serios y desarrollo de software.

- Describir antecedentes de metodologías utilizadas en el diseño de software, en particular de juegos serios, así como las arquitecturas de software y de despliegue utilizadas.

- Elaborar una propuesta metodológica y arquitectónica para el diseño de juegos serios que potencien la adquisición de competencias tecnológicas.

- Validar la metodología a través de un panel de expertos o en el desarrollo de un prototipo no funcional de juego serio, de manera que permita analizar el alcance del aporte de esta tesis y las posibles mejoras por realizar.

- Elaborar un informe con los resultados y las conclusiones a las que se arriba.

## Aportes de la tesis

El principal aporte que brinda esta investigación radica en una propuesta metodológica y arquitectónica para orientar y guiar el proceso de diseño e implementación de prototipos de juegos serios, específicamente, los que han sido pensados para ser utilizados en la formación de habilidades o competencias tecnológicas.

Otro de los aportes importantes de la tesis consiste en que, a partir de la metodología desarrollada, permite identificar cual es la población meta más apta para utilizar el juego serio en la formación de habilidades o competencias tecnológicas.

## Líneas de I/D futuras

Se trabajará en el diseño e implementación del prototipo funcional del juego serio **SkillNet**, se considerará la posibilidad de que el mismo sea ofrecido como un software de servicio. Se debe trabajar en la aplicación de la propuesta metodológica **INTEGRA+506** para el diseño de prototipos de juegos serios en diferentes ámbitos educativos, así como en distintas áreas del conocimiento y competencias.



**Advisor**  
Dra. Patricia Alejandra Bazán

**Thesis approval date**  
June, 15 2023

**SEDICI**  
[Full Text](#)

## **Methodological and architectural proposal to develop serious games oriented to the formation of technological skills**

**Keywords:** Serious games; technological skills training; prototype; teaching staff; methodological and architectural proposal; INTEGRA+506; SkillNet; data networks.

### **Motivation**

Given the emergence of Information and Communication Technologies (ICTs) and their rapid progress and impact at the educational level, Higher Education Institutions (HEIs) are forced to face the challenge of keeping up to date with the new demands of today's society (Vega-Hernández et al., 2018). Likewise, (Cabero et al., 2015) point out that ICTs provide HEIs with the potential for the creation of learning scenarios that allow immediate access to both information and technological resources that facilitate the ongoing training of students and faculty in digital technologies (It is essential to clarify that for cultural reasons the word "teacher" is used in Costa Rica as a synonym for "professor" or "faculty". Therefore, both terms are used synonymously in this document). Aguaded & Cabero (2014) state that the main problem expressed by professors is professional updating since they require various training activities related to enhancing their technological skills and competencies, which are required to use and integrate ICT in the educational processes.

In research conducted by Teo (2009) and Scherer et al. (2015), it is stated that ICT's teaching and learning process can be innovated by incorporating serious games in the work methodologies. This is because they can be used for the acquisition of skills, formation of competencies, changes in attitude and behavior, and enhancing motivation and learning performance. In this line, serious games can become a medium for teachers to approach digital technologies (Boyle et al., 2011; Chang et al., 2012).

Sandí & Bazán (2017) conducted research to pose challenges and difficulties for the execution of educational serious games using the possibilities provided by the SaaS model in Cloud Computing. As a result of the work, a list of functional characteristics that educational serious games would acquire when used as software as a service was obtained. In this same line, Sandí & Sanz (2018) carried out research

to investigate the state of the art of educational serious games and their possibilities for acquiring technological competencies in teacher training. As a result, a proposal for integrating serious games to form technological competencies in teachers and improve their attitudes towards these technologies was obtained. In addition, a proposal was elaborated to evaluate the professors' degree of acceptance of digital technologies, mainly on serious games.

Although there is a variety of research that refers to the possibility offered by serious games for developing digital skills and competencies, there is no evidence of serious games designed exclusively to enhance teachers' technological skills and competencies at the higher education level. Likewise, it has yet to be located or identified in the bibliographic references analyzed the existence of specific methodologies that orient or serve as a guide for the creation and design of serious games oriented to the formation of digital competencies.

The results obtained in the research carried out give rise to a theoretical reference for the elaboration of research that allows proposing a methodology and software architecture that guides the steps to follow during the creation and implementation of serious games that are going to be used for the training of technological skills or competencies. Consequently, it is proposed to apply the proposed methodology and software architecture to create and design a non-functional prototype of a serious game that can be used for the training of technological skills or competencies.

#### **a) General objective**

- To define a methodology and software architecture for the design of serious games that allow the acquisition or training of technological competencies.

#### **b) Specific objectives**

- To study the basic concepts related to ICT, technological skills, serious games and software development.

- To describe the background of methodologies used in software design, in particular for serious games, as well as the software and deployment architectures used.
- To elaborate a methodological and architectural proposal for the design of serious games that promote the acquisition of technological competencies.
- To validate the methodology through a panel of experts and the development of a nonfunctional prototype of a serious game, in order to analyze the scope of the contribution of this thesis and the possible improvements to be made.
- To elaborate a report with the results and the conclusions reached

### **Thesis contributions**

The main contribution of this research lies in a methodological and architectural proposal to orient and guide the design and implementation process of serious game prototypes, specifically, those designed to be used in the training of technological skills and/or competencies.

Another significant contribution of the thesis is that based on the methodology developed, and it is possible to identify the target population most suitable for using the serious game in training technological skills or competencies.

### **Future Research Lines**

A working prototype of the SkillNet serious game will be designed and implemented. Also, this prototype will be considered to be offered as a service software.

Work should be done on applying the INTEGRA+506 methodological proposal for the design of prototypes of serious games in different educational environments and in different areas of knowledge and competencies.



español

**Directores**  
Lic. Patricia Pesado  
Mg. Rodolfo Bertone

**Asesor Científico**  
Dr. Hernán Merlino

**Fecha de aprobación**  
26 de junio de 2023

**SEDICI**  
[Tesis Completa](#)

## Integración de métodos de descubrimiento de conocimiento embebido en fuentes de información desestructuradas

**Palabras clave:** procesamiento de lenguaje natural; PLN; Open Information Extraction, Open IE; relaciones semánticas; extracción de información

### Motivación

El objetivo o motivación de esta tesis fue la de construir una familia de métodos de extracción de conocimiento, tal que, dado un documento como entrada con información no estructurada pero con conocimiento embebido (por ejemplo un texto en lenguaje natural), sean capaces de generar un conjunto de piezas de conocimiento estructurado, por ejemplo: reglas de producción, subgráficos de una red semántica, etcetera.

Objetivos específicos fueron: construir un método de Open IE para idioma español en el estado del arte, armar un marco de referencia para evaluar diversos métodos existentes y proponer un método novedoso de OpenIE.

### Aportes de la tesis

#### Creación de métodos de Open IE para idioma español.

Se crearon dos métodos de extracción de conocimiento para la Web en idioma español: TP-OIE-ES y ECMes. Se puede concluir que ECMes es, hoy por hoy, 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.

#### Creación de un marco de referencia para la evaluación de los métodos de Open IE.

Se puede afirmar que se cumplió el objetivo secundario propuesto con la creación de dos conjuntos de datos para la evaluación de métodos de extracción de conocimiento para la Web, uno en inglés: Reuters-103 y otro en español: conjunto Rodríguez. Ambos conjuntos contienen textos en lenguaje natural y para cada oración o párrafo tienen una serie de relaciones semánticas creadas a mano.

#### Objetivo secundario: creación de un método de Open IE novedoso.

Para cumplir con este objetivo se creó primeramente el método TP-OIE, el cual sirvió para poner a prueba la mecánica propuesta. A partir de los prometedores resultados de ATP-OIE, se diseñaron los métodos: TP-OIE-ES y ECMes para idioma español.

### Líneas de I/D futuras

Todo método de extracción de conocimiento para la Web debe ser comparado, en última instancia, con la capacidad humana de comprender y analizar un texto. Estos métodos permiten, de forma automática y mediante algún procesamiento posterior, analizar, resumir o incluso crear mapas de conocimiento de textos en lenguaje natural. Son utilizados cuando el volumen de datos a procesar, el corpus, es demasiado grande para poder ser analizado por personas. Los errores: falta de precisión, falta de exhaustividad, relaciones semánticas poco informativas y manejo de información subjetiva, son todos errores debidos a la misma causa raíz: falta de entendimiento o de comprensión de un texto.

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 (2020) y la otra es BERT (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 analizados en el documento de tesis: Multi2OIE, 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 comprensión de textos de forma automática. (junio de 2021)



english

**Advisors**

Lic. Patricia Pesado  
Mg. Rodolfo Bertone

**Scientist Consultant**

Dr. Hernán Merlino

**Thesis approval date**

June 26, 2023

**SEDICI**

[Full Text](#)

## Integration of Embedded Knowledge Discovery Methods in Unstructured Information Sources

**Keywords:** natural language processing; PLN; Open Information Extraction, Open IE; semantic relationships; information extraction

### Motivation

The objective or motivation of this thesis was to build a family of knowledge extraction methods, such that, given a document as input with unstructured information but embedded knowledge (for example, a text in natural language), they are capable of generating a set of structured pieces of knowledge, such as production rules, subgraphs of a semantic network, etc.

Specific objectives were: to construct a state-of-the-art Open IE method for the Spanish language, establish a reference framework for evaluating various existing methods, and propose a novel Open IE method.

### Thesis contributions

#### Creation of Open IE Methods for the Spanish Language

Two knowledge extraction methods for the web in the Spanish language were developed: TP-OIE-ES and ECMes. It can be concluded that ECMes is, currently, a state-of-the-art knowledge extraction method for the web in the Spanish language and performs the best among the test sets used.

#### Creation of a Framework for Evaluating Open IE Methods

It can be stated that the secondary objective proposed was achieved with the creation of two datasets for the evaluation of knowledge extraction methods for the web, one in English: Reuters-103, and another in Spanish: the Rodríguez dataset. Both datasets contain texts in natural language and a set of manually created semantic relations associated with each example.

#### Secondary Objective: Creation of a Novel Open IE Method

To fulfill this objective, the TP-OIE method was initially created, which served to test the proposed mechanics. Based on the promising results of ATP-OIE, the methods TP-OIE-ES and ECMes were designed for the Spanish language

### Future Research Lines

Every knowledge extraction method for the web must ultimately be compared to the human ability to comprehend and analyze a text. These methods allow, automatically and through some subsequent processing, the analysis, summarization, or even the creation of knowledge maps from natural language texts. They are used when the volume of data to be processed, the corpus, is too large to be analyzed by humans. Errors such as lack of precision, lack of completeness, uninformative semantic relationships, and handling of subjective information are all errors stemming from the same root cause: a lack of understanding or comprehension of a text.

In parallel with the proliferation of knowledge extraction methods for the web, some specific models have emerged in an attempt to create artificial intelligence capable of “understanding” natural language text or coming as close as possible to this goal. In particular, these are pre-trained artificial neural network models. One of them is GPT-3 (2020), and the other is BERT (2019). Despite the promise of these models, they do not achieve natural language processing at the same level as a human being. Even one of the methods analyzed in the thesis document: Multi2OIE, internally uses BERT and yet is not the best-performing method.

If the goal of building an AI capable of understanding natural language text in the same way a human does is achieved, knowledge extraction methods for the web would become obsolete, as their functionality would be just one of the many feasible tasks for these systems. Therefore, the problems posed will remain open until automatic text comprehension can be satisfactorily resolved. (June 2021).



**Directora**  
Dra. Alejandra Garrido

**Codirectores**  
Dr. Julián Grigera  
Dr. Gustavo Héctor Rossi

**Fecha de aprobación**  
28 de junio de 2023

**SEDICI**  
[Tesis Completa](#)

## **Métodos y herramientas de desarrollo y evaluación de refactorings para la mejora de la experiencia de usuario en aplicaciones web**

**Palabras clave:** Experiencia de Usuario; Interacción de usuario; Métodos Ágiles, Refactorings de UX; Aprendizaje Automático; End-user Development.

### **Motivación**

La experiencia de usuario (UX) es un factor crucial que determina el éxito de las aplicaciones web. Si bien actualmente se reconoce la importancia de proveer una experiencia de usuario óptima, las prácticas de evaluación y mejora de la UX suelen postergarse. Más allá de los recursos necesarios como expertos en UX y usuarios para hacer pruebas, el tiempo requerido por estas prácticas es el principal motivo por el cual resulta difícil integrarlas en los métodos ágiles ampliamente utilizados hoy en día.

Para facilitar la evaluación de la UX, se han desarrollado métodos automáticos para detectar y solucionar problemas de interacción. Uno de los métodos propuestos es la técnica de refactoring de usabilidad, la cual se basa en aplicar pequeñas transformaciones (refactorings) a través scripts ejecutados en el navegador que sin alterar la funcionalidad de la aplicación pretenden mejorar la UX. Esta técnica fue propuesta como una solución automática, en la que los refactorings aparecen como soluciones definitivas sin antes poder validar cómo impactan en los usuarios.

En esta tesis se desarrolla un conjunto de métodos y herramientas para dar soporte en las tareas de evaluación y mejora de la UX. En primer lugar se presenta un método para explorar alternativas de diseño haciendo uso de la técnica de refactoring. Este método se desarrolla en una herramienta llamada UX-Painter, mediante la cual un diseñador puede crear distintas versiones de una aplicación web utilizando los refactorings para generar los cambios de diseño, sin la necesidad de codificarlos. De esta manera, es posible visualizar, validar y evaluar cambios de diseño en la misma aplicación web sin que éstos tengan que ser implementados. Respecto a la evaluación de los diseños, no solo es impor-

tante determinar el impacto de un cambio de diseño en la experiencia del usuario, sino que además resulta fundamental poder comparar distintas alternativas. Sobre todo porque los refactorings en ciertos casos ofrecen más de una solución posible para un mismo problema. Por este motivo, en esta tesis también se desarrolla una métrica denominada esfuerzo de interacción que permite evaluar y comparar diseños alternativos en páginas web. Se trata de un puntaje asignado por expertos en UX que asigna sobre los distintos widgets de una página web, que son aquellos elementos que se ven modificados por los refactorings. Para calcularlo automáticamente, se desarrollaron modelos de predicción que toman como entrada diferentes logs de interacción de usuario que reciben el nombre de micro-medidas. El hecho que la métrica sea transparente a los usuarios hace que pueda ser utilizada para evaluar diseños con múltiples usuarios en experimentos online como A/B testing o similares. Como prueba de concepto, se implementa UX-Analyzer, una herramienta web que permite calcular y visualizar el esfuerzo de interacción de distintas versiones de una aplicación web.

Finalmente, una vez que se verifica que ciertos refactorings mejoran la UX, deben ser implementados en la aplicación bajo análisis. Para facilitar esta tarea a los desarrolladores, se presenta un mecanismo para generar una implementación preliminar de los refactorings usando como base los frameworks y librerías utilizados actualmente para el desarrollo de interfaces web.

El trabajo incluye distintas validaciones que comprueban la aplicabilidad de los desarrollos propuestos para facilitar la mejora de la UX.

## Aportes de la tesis

Un método de End-user Development para explorar variantes de diseño utilizando la técnica de refactoring.

La incorporación de nuevos Client-side web refactorings al catálogo desarrollado en trabajos anteriores.

Un mecanismo de adaptación de estilos que permite personalizar ciertas propiedades estéticas de los cambios de diseño impuestos por los refactorings.

Una herramienta denominada UX-Painter que implementa los 3 puntos anteriores.

La métrica llamada esfuerzo de interacción que sirve para evaluar el funcionamiento de los distintos elementos o widgets de la UI que se ven modificados por los refactorings.

Un conjunto de micro-medidas para cada uno de los 6 tipos de widget considerados por la métrica anterior que cuantifican diferentes aspectos de la interacción del usuario.

Modelos de predicción que permiten estimar el esfuerzo de interacción de un widget utilizando como entrada las micro-medidas calculadas automáticamente.

Una herramienta llamada UX-Analyzer que utiliza los modelos anteriores para calcular y visualizar el esfuerzo de interacción de distintas versiones de una página web.

## Líneas de I/D futuras

- Desarrollar nuevos Client-side web refactorings para contar con más opciones para explorar cambios de diseño.
- Continuar con el desarrollo del mecanismo para generar automáticamente una implementación preliminar de los refactorings.
- Realizar validaciones de las herramientas UX-Painter y UX-Analyzer en un contexto de uso real para determinar empíricamente su aplicabilidad y potenciales mejoras.
- Incorporar nuevos tipos de widget a la métrica de esfuerzo de interacción.
- Agregar a UX-Analyzer otro tipo de métricas que también contribuyan a la evaluación de la UX.



**Advisor**

Dra. Alejandra Garrido

**Codirectors**

Dr. Julián Grigera

Dr. Gustavo Héctor Rossi

**Thesis approval date**

June 28, 2023

**SEDICI**

[Full Text](#)

## **Methods and tools for refactoring development and evaluation to improve users experience in web applications**

**Keywords:** User Experience; User Interaction; Agile Methods, UX Refactorings; Machine Learning; End-user Development

### **Motivation**

User Experience (UX) is a crucial aspect that determines the success of web applications. Although companies recognize the importance of providing a good user experience, UX evaluation is often neglected. Besides the resources needed such as experts and users, UX practices are time-consuming, which makes it difficult to integrate them with agile methods. With the goal of facilitating the UX evaluation, different methods have been proposed to automatically detect and fix user interaction problems. One of these methods is the usability refactoring technique, which performs small transformations (known as refactorings) to the user interface through predefined scripts executed on the browser. However, while refactorings provide an automatic way for UX improvement, they are applied without assessing the real gain on the user experience of the target application.

In this thesis we develop a set of methods and tools to support UX experts on the evaluation and improvement of the UX. First, we present a method to explore design alternatives using the refactoring technique. This method is developed in a tool called UX-Painter, which can be used by a UX designer to create alternative versions of a web application by applying the refactorings to generate the design changes, without having to code them. In this way, designers can visualize, validate and evaluate design changes on the web application before implementing them.

Regarding the design evaluation, it is important not only to assess the impact of each design change, but also to compare different alternatives. Given that in some cases the refactorings offer more than one solution for a specific problem, in this work we develop a metric called interaction effort that allows to evaluate and compare alternative designs on web pages. This metric is a score assigned by UX experts on

the widgets of a web page, which are the elements modified by the refactorings. In order to automatically calculate the interaction effort, we develop machine learning models that predict the score using user interaction logs called micro-measures. The metric is transparent to the users, which makes it suitable for evaluating designs in online experiments such as A/B testing. As a proof of concept, we also show UX-Analyzer, a web tool to calculate and visualize the interaction effort of different versions of a web application.

Finally, after validating that a set of refactorings effectively improve the UX, they have to be implemented in the target application. With the aim of facilitating this task, we present a mechanism to automatically generate a preliminary implementation of the refactorings. This implementation is based on the frameworks currently used to build user interfaces on the web.

This work includes different validations that assess the applicability of the approach proposed to facilitate the UX improvement.

### **Thesis contributions**

An End-user Development method to explore alternative designs using the refactoring technique.

New Client-side web refactorings that were added to the catalogue defined in previous works.

A style adaptation mechanism that allows to personalize certain aesthetics properties of the design changes performed by the refactorings.

A tool called UX-Painter that implements the previous items.

A metric named interaction effort to assess the UX of the different UI widgets modified by the refactorings.

A set of micro-measures for each of the 6 widget types



considered in the interaction effort that quantify different aspects of the user interaction.

Prediction models to estimate the user interaction effort with a widget using as input the micro-measures automatically calculated.

UX-Analyzer, a web tool that uses the prediction models to calculate and visualize the interaction effort of different versions of a web application.

### **Future Research Lines**

- Develop new Client-side web refactorings to have more options to explore alternatives designs.
- Continue with the development of the mechanism to automatically generate an implementation for the refactorings.
- Validate the tools UX-Painter and UX-Analyzer in a real context of use to empirically determine their applicability and to identify potential improvements.
- Incorporate new widget types to the interaction effort metric.
- Add to UX-Analyzer others metrics that contribute to the UX evaluation.



Director  
Ing. Armando De Giusti

Fecha de aprobación  
13 de julio de 2023

Repositorio SEDICI  
[Tesis Completa](#)

## **Análisis de datos educativos aplicado en el estudio de la incidencia de factores socioeconómicos en el rendimiento escolar**

**Palabras clave:** aprendizaje automático; sobremuestreo; Orange; minería de datos educativos; análisis confirmatorio; SHAP

### **Motivación**

Esta investigación se desarrolló en el campo de la Minería de Datos Educativos. Articula el análisis de datos con el problema multifactorial del rendimiento académico escolar motivado por la incidencia de factores socioeconómicos. Se han aplicado modelos supervisados, no supervisados y un análisis confirmatorio sustentado en exploraciones estadísticas que preceden a los modelos.

Para muchos niños, dejar la escuela es el final de un proceso de desconexión gradual y de reducción de su participación en el currículo formal y la vida social escolar. La deserción desencadena mayormente en la reprobación de materias a causa de un insuficiente rendimiento, que según Tarik y colaboradores (2021), puede suscitar que varios estudiantes adopten comportamientos antisociales o luego situarse en la franja laboral más baja.

Según la UNESCO (2017), las escuelas empiezan a concebir al análisis de datos como oportunidades educativas significativas para reducir la deserción bajo la premisa que cuando se sepa de estudiantes menos favorecidos se tendrá probabilidades de mejorar su estancia educativa.

La minería de datos educativos favorece la búsqueda de relaciones entre los puntajes de los alumnos, exploración de conocimientos sobre el aprendizaje contextualizado, análisis de factores socioeconómicos que lo afectan de modo clave o tendencias de desarrollo a partir del rendimiento y comportamiento actual (McNeish & Wolf, 2020).

El sistema escolar ecuatoriano consta de tres años enfocados en el fortalecimiento cognitivo, afectivo y psicomotriz; y tres años que abordan destrezas y técnicas de estudio. Los datos analizados incluyen promedios quimestrales y

parciales, además, datos socioeconómicos como: estructura familiar, ingreso familiar, parentesco del representante, dirección/distancia domiciliaria, cantidad de personas que viven en casa, escolaridad alcanzada por padres y disponibilidad de servicios básicos como acceso a computadoras e internet.

La pandemia del COVID-19 reveló fragilidades en los sistemas educativos, a octubre de 2020, según la UNESCO (2021), un tercio de los estudiantes del mundo no tenían acceso a la educación online debido a la falta de los dispositivos o conectividad, o sus escasas habilidades digitales.

Entonces: ¿Cómo inciden los factores socioeconómicos en el aprovechamiento escolar?

#### **Objetivos**

**General:** Analizar la incidencia de los factores socioeconómicos en el aprovechamiento académico a nivel escolar y de ese modo contribuir a su entendimiento y mejora, mediante la aplicación de modelos de análisis de datos.

#### **Específicos:**

1. Reconocer las aplicaciones de análisis de datos en los problemas del contexto educativo escolar.
2. Preparar los datos acordes con la dimensionalidad a un número efectivo de características.
3. Estudiar comparativamente algoritmos de minería de datos en función del proceso de minería de datos que resulte idóneo.
4. Establecer parámetros e hiperparámetros apropiados a los datos y los modelos.
5. Interpretar los resultados del conocimiento descubierto y su eficiencia según métricas pertinentes a los modelos.

## Aportes de la tesis

1. Contribución al área de análisis de datos para el mejoramiento educativo, teniendo en cuenta información disponible en Ecuador (área donde se desempeña el Tesista), pero con resultados proyectables en forma general debido al impacto de la socioeconomía en el rendimiento académico.
2. Análisis de datos académico y socioeconómicos escolares.
3. Revisión de calificaciones finales y de avances progresivos para generar recomendaciones tempranas sustentadas en datos que disipen posibles afectaciones al aprovechamiento.
4. Exploración y configuración los datos con criterios de ingeniería, con hipótesis y sus validaciones en tareas de clasificación y predicción, hasta evidenciar métricas favorables que sugieran resultados interpretables y susceptibles de convertirse en recomendaciones para una apropiación social que beneficie a la educación escolar.

## Líneas de I/D futuras

Se aspira a investigar para sustentar acciones respaldadas con datos e Inteligencia Artificial que ayuden a los niños a una estancia escolar más agradable:

- Incorporación de más escuelas de distintas regiones para obtener un abordaje más significativo.
- Revisiones literarias en más bases de datos e idiomas empleando inteligencia artificial para conseguir cantidad y calidad literaria.
- Observación de cada factor socioeconómico y el favorecimiento de medidas proactivas que mitiguen la problemática multifactorial.
- Generación de soluciones de aprendizaje automático interactivas e interpretables con base en normativas emergentes que transparenten a los usuarios la razón de sus decisiones.
- Abordaje de datos psicológicos, estilos de aprendizaje, autoeficacia, ambiente hogareño, motivación u otra información obtenible mediante encuestas periódicas que evidencien su variabilidad situacional temporal. Los psicólogos de las escuelas usualmente recurren a textos (analizables con minería de texto) manuscritos por los alumnos que pueden ser enriquecidos con sus anotaciones para contrastar habilidades motoras finas, lenguaje, memoria y concentración, con su rendimiento.



Advisor  
Ing. Armando De Giusti

Thesis approval date  
July 13, 2023

SEDICI  
[Full Text](#)

## Analysis of educational data applied in the study of the incidence of socioeconomic factors in school performance

**Keywords:** machine learning; oversampling; Orange; educational data mining; confirmatory analysis; SHAP.

### Motivation

This research was developed in the field of Educational Data Mining. It articulates the data analysis with the multifactorial problem of school academic performance motivated by the incidence of socioeconomic factors. Supervised and unsupervised models and a confirmatory analysis supported by statistical explorations that precede the models have been applied.

For many children, leaving school is the end of a process of gradual disengagement and reduced participation in the formal curriculum and school social life. Dropout mainly triggers the failure of subjects due to insufficient performance, which according to Tarik et al. (2021), can lead several students to adopt antisocial behaviors or later be placed in the lowest labor class.

According to UNESCO (2017), schools begin to conceive of data analysis as a significant educational opportunity to reduce dropout under the premise that when less favored students are known, it will be likely to improve their educational stay.

Educational data mining favors the search for relationships between student scores, exploration of knowledge about contextualized learning, analysis of socioeconomic factors that affect it in a key way, or development trends based on current performance and behavior (McNeish & Wolf, 2020).

The Ecuadorian school system consists of three years focused on cognitive, affective, and psychomotor strengthening; and three years that address study skills and techniques. The data analyzed includes quarterly

and partial averages, in addition, socioeconomic data such as family structure, family income, the relationship of the representative, address/distance from home, number of people living at home, schooling achieved by parents, and availability of basic services such as access to computers and internet.

The COVID-19 pandemic revealed weaknesses in educational systems, as of October 2020, according to UNESCO (2021), a third of the world's students did not have access to online education due to a lack of devices or connectivity, or their scant digital skills.

So: How do socioeconomic factors affect school achievement?

#### Purpose

**General:** Analyze the incidence of socioeconomic factors in academic achievement at the school level and thus contribute to its understanding and improvement, through the application of data analysis models.

#### Specifics:

1. Recognize the applications of data analysis in the problems of the school educational context.
2. Prepare the data according to the dimensionality of an effective number of characteristics.
3. Comparatively study data mining algorithms based on the ideal data mining process.
4. Set appropriate parameters and hyperparameters to the data and models.
5. Interpret the results of the discovered knowledge and its efficiency according to metrics relevant to the models.

## Thesis contributions

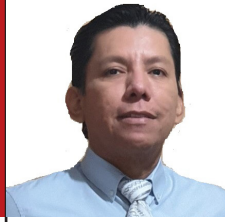
1. Contribution to the area of data analysis for educational improvement, considering information available in Ecuador (the area where the Thesis Student works), but with generally projectable results due to the impact of socioeconomics on academic performance.
2. Analysis of academic and socioeconomic school data.
3. Review of final qualifications and progressive advances to generate early recommendations supported by data that dispel possible effects on the use.
4. Exploration and configuration of the data with engineering criteria, with hypotheses and their validations in classification and prediction tasks, until favorable metrics are evidenced that suggest interpretable results and capable of becoming recommendations for a social appropriation that benefits school education.

## Future Research Lines

Research is aspired to support actions supported with data and Artificial Intelligence that help children to have a more pleasant school stay:

1. Incorporation of more schools from different regions to obtain a more significant approach.
2. Literary reviews in more databases and languages using artificial intelligence to achieve literary quantity and quality.
3. Observation of each socioeconomic factor and the favoring of proactive measures that mitigate multifactorial problems.
4. Generation of interactive and interpretable machine learning solutions based on emerging regulations that make the reason for their decisions transparent to users.
5. Approach to psychological data, learning styles, self-efficacy, home environment, motivation, or other information obtainable through periodic surveys that demonstrate their temporary situational variability. School psychologists usually resort to texts (analyzable with text mining) handwritten by students that can be enriched with their annotations to compare fine motor skills, language, memory, and concentration, with their performance.

## DOCTORADO EN CIENCIAS INFORMÁTICAS



**Dr. Luis Orlando Philco Asqui**

e-mail

orlandophilco\_7@hotmail.com

español

**Director**

Ing. Luis Marrone

**Fecha de aprobación**

5 de septiembre de 2023

**SEDICI**

[Tesis Completa](#)

## **Calidad de sensado y eficiencia energética en redes de sensores inalámbricos (WSN)**

**Palabras clave:** IoT; WSN; Calidad de Sensado; Eficiencia Energética; 4TQS; MiA-CODER

### **Motivación**

Para este autor el diseño de estrategias de calidad de sensado es la base motivacional para el problema de agotamiento prematuro de la energía en los nodos de una red de sensores

inalámbricos (WSN), Así se originan en una región determinada la división de la red o creación de agujeros de cobertura. Se indican dos objetivos principales: Diseñar y evaluar método basado en cuatro fases para calidad de sensado en red estrella y malla y Diseñar y evaluar algoritmos para calidad de sensado basada en aprendizaje por refuerzo con multiagentes inteligentes para topología árbol de clústeres desiguales.

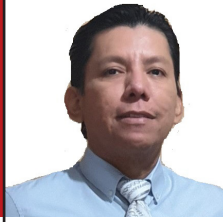
### **Aportes de la tesis**

Dos mecanismos que detecten y reparen agujeros de cobertura para evitar derroche de energía o elevado costo energético en una red de sensores inalámbricos. Además de asegurar conectividad equilibrada donde se reduzcan pérdida de paquetes en la red que sea desplegada de forma realista (aleatorio) tanto en topología estrella, malla y de tipo árbol.

### **Líneas de I/D futuras**

Inteligencia artificial en ciberseguridad de WSN, Sistemas de sensores y actuadores para entornos industriales. Sistemas de sensores acústicos submarinos.

DOCTORADO EN  
CIENCIAS INFORMÁTICAS



**Dr. Luis Orlando Philco Asqui**

e-mail

orlandophilco\_7@hotmail.com

english

**Advisor**  
Ing. Luis Marrone  
**Thesis approval date**  
September 5, 2023  
**SEDICI**  
[Full Text](#)

## **Quality of sensing and energy efficiency in Wireless Sensor Networks (WSN)**

**Keywords:** IoT; WSN; Quality of Sensing; Energy efficiency; 4TQS; MiA-CODER

### **Motivation**

For this author, the design of sensing quality strategies is the motivational basis for the problem of premature depletion of energy in the nodes of a wireless sensor network (WSN). Thus, the division of the network or creation of cover holes. Two main objectives are indicated: Design and evaluate a method based on four phases for quality sensing in star and mesh networks and Design and evaluate algorithms for quality sensing based on reinforcement learning with intelligent multiagents for unequal cluster tree topology.

### **Thesis contributions**

Two mechanisms that detect and repair coverage holes to avoid energy waste or high energy cost in a wireless sensor network. In addition to ensuring balanced connectivity where packet loss is reduced in the network that is deployed in a realistic (random) way in both star, mesh and tree-type topologies.

### **Future Research Lines**

Artificial intelligence in cybersecurity from WSN, Sensor and actuator systems for industrial environments. Underwater acoustic sensor systems.



**Director**  
Lic. Javier Díaz

**Fecha de aprobación**  
11 de septiembre de 2023

**SEDICI**  
[Tesis Completa](#)

## **Método de auditoría Informática basado en sistemas de procesamiento avanzado de datos que permita minimizar el riesgo de calidad de los resultados**

**Palabras clave:** Auditoría informática; Instituciones de Educación Superior; calidad; seguridad; estándar; métricas

### **Motivación**

Las Instituciones de Educación Superior del Ecuador no cuentan con marcos de referencia regulatorios ni con muchos que permitan a los auditores informáticos emitir criterios objetivos al realizar evaluaciones integrales para determinar la situación actual de la seguridad de la información generada en los procesos académicos universitarios, ya que los marcos establecidos están enfocados en el ámbito empresarial, se abre de esta forma una brecha importante en el estudio de la auditoría informática en seguridad de la información, tomando en cuenta que los auditores realizan estas actividades a su mejor criterio y en algunos casos auditores tienen escasos conocimientos de nuevas técnicas de tratamiento de datos y utilizan metodologías tradicionales.

#### **Objetivos:**

**General:** Diseñar un método de auditoría informática, basado en la estandarización de procesos, que permita minimizar el riesgo de la calidad y seguridad de la información obtenida de IES.

#### **Específicos:**

- Describir la calidad de los resultados que obtienen los auditores en las instituciones de educación superior de la Zona 1 de Ecuador, durante los procesos de auditoría informática.
- Determinar el nivel de seguridad con el que cuenta la información obtenida por los auditores en las instituciones de educación superior de la Zona 1 de Ecuador, durante los procesos de auditoría.
- Diagnosticar el grado de estandarización del procedimiento de auditoría que llevan a cabo los auditores en las instituciones de educación superior, para obtener la información.
- Analizar las alternativas de auditoría informática existentes, en términos de los criterios de estandarización de la auditoría
- Comparar las alternativas de auditoría informática exist-

tentes, en términos de los criterios de estandarización de la auditoría.

- Explicar cómo el proceso de estandarización de la auditoría afecta la calidad y la seguridad de los resultados obtenidos por los auditores en las instituciones de educación superior.
- Proponer un método de auditoría informática basado en sistemas de procesamiento avanzado de datos que permita minimizar el riesgo de calidad y de seguridad de los resultados.

### **Aportes de la tesis**

El aporte principal de este trabajo investigativo es el diseño de un método de auditoría en seguridad de la información que utilice técnicas avanzadas de tratamiento de datos, para asegurar la calidad, objetividad y confiabilidad de los resultados. La utilización de técnicas de auditoría informática contribuirá al desarrollo del método con buenas prácticas recomendadas por marcos de referencia internacionales.

### **Líneas de I/D futuras**

Auditoría Informática utilizando Inteligencia Artificial como mecanismo de minimizar el riesgo de subjetividad.





english

**Advisor**  
Lic. Javier Díaz  
**Thesis approval date**  
September 11, 2023  
**SEDICI**  
[Full Text](#)

## **Computer Audit Method Based on Advanced Data Processing Systems to Minimize the Risk of quality issues**

**Keywords:** Computer audit; Higher Education Institutions; quality; security; standards; metrics

### **Motivation**

Higher Education Institutions in Ecuador lack regulatory frameworks and methods that enable IT auditors to provide objective criteria when conducting comprehensive assessments to determine the current state of information security generated in university academic processes. This gap exists because the established frameworks are primarily focused on the business sector. Consequently, there is a significant deficiency in the study of computer audits in information security, considering that auditors carry out these activities based on their best judgment, and in some cases, auditors have limited knowledge of new data processing techniques and rely on traditional methodologies.

#### **Objectives:**

**General:** To design a computer audit method based on process standardization that minimizes risks related to the quality and security of information obtained from Higher Education Institutions (HEIs).

#### **Specifics:**

- To describe the quality of the results obtained by auditors in higher education institutions in Zone 1 of Ecuador during the processes of computer auditing.
- To determine the level of security associated with the information obtained by auditors in higher education institutions in Zone 1 of Ecuador during the auditing processes.
- To diagnose the degree of standardization of the auditing procedure carried out by auditors in higher education institutions to obtain information.
- To analyze the existing alternatives for computer auditing in terms of audit standardization criteria.
- To compare the existing alternatives for computer

auditing in terms of audit standardization criteria.

- To explain how the standardization process of auditing affects the quality and security of the results obtained by auditors in higher education institutions.
- To propose a computer audit method based on advanced data processing systems that minimize quality and security risks in the results.

### **Thesis contributions**

The main contribution of this research work is the design of an information security audit method that utilizes advanced data processing techniques to ensure the results' quality, objectivity, and reliability. The use of computer audit techniques will contribute to the development of the method with best practices recommended by international reference frameworks.

### **Future Research Lines**

Computer Auditing using Artificial Intelligence as a mechanism to minimize subjectivity risk.



**Directora**  
Dra. Sandra Baldassarri

**Codirector**  
Mg. Hugo Ramon

**Fecha de aprobación**  
27 de septiembre de 2023

**SEDICI**  
[Tesis Completa](#)

## **Modelado e implementación de algoritmos inteligentes e análisis de opinión**

**Palabras clave:** Análisis de Sentimientos; Procesamiento del Lenguaje Natural; Supervisión Distante; Aprendizaje Automático; Información Contextual

### **Motivación**

A la par de la amplia adopción que han tenido las redes sociales, ha crecido también la generación de contenidos en ellas, en particular en forma de texto. La proliferación de este tipo de contenido ha creado la materia prima necesaria para aplicar técnicas de minería de textos a esos datos con el objetivo de extraer información valiosa.

Numerosos trabajos que intentan categorizar, mediante clasificadores basados en aprendizaje automático, textos provenientes de redes sociales, dependen del etiquetado manual del contenido o de la utilización de datasets públicos previamente etiquetados. Dichos abordajes presentan sus inconvenientes, uno de ellos es el tiempo que demanda la clasificación manual de los datos de entrenamiento. Otro problema es que los clasificadores suelen construirse utilizando datos de distinto origen a los que finalmente analizan, esto plantea un desafío debido a que, si el clasificador no fue expuesto, durante la etapa de entrenamiento, a datos similares a los que finalmente debe categorizar, difícilmente pueda hacerlo de manera adecuada. Por otro lado, la cantidad de recursos disponibles (tales como datasets etiquetados, corpus o diccionarios afectivos) no es abundante para idiomas distintos del inglés, limitando las posibilidades de construcción de los mencionados clasificadores de texto para otros idiomas, entre ellos el español.

La tarea de recopilación y validación de recursos en el idioma a utilizar se vuelve, en consecuencia, una necesidad para construir clasificadores de texto, basados en aprendizaje automático supervisado. Sin embargo, dichas tareas son extremadamente demandantes en tiempo y recursos humanos. Esta problemática se agrava para los casos en

los que el criterio de clasificación no es objetivo, como por ejemplo para la clasificación de emociones en texto. En estas situaciones, se requiere que varios jueces clasifiquen el mismo contenido, de manera de poder validar la veracidad de la etiqueta asignada al mismo.

Para poder agilizar el desarrollo de clasificadores de emociones en texto para el idioma español basados en aprendizaje automático supervisado, resulta necesario reducir o eliminar la necesidad del etiquetado manual de los datasets utilizados para entrenamiento. Para ello se establecen los siguientes objetivos específicos:

- Construir o adquirir conjuntos de datos, generados a partir de la interacción en redes sociales, que reflejen la opinión del público en diversos sucesos y en lo posible incluyan información contextual acerca de los sucesos previamente mencionados.
- Realizar tareas de limpieza y pre procesamiento sobre los textos de entrada, con el objetivo de verificar su impacto en el desempeño de clasificadores basados en aprendizaje automático.
- Definir una metodología y seleccionar una métrica apropiada para validar las etiquetas asignadas a los textos recopilados.
- Implementar clasificadores de texto basados en aprendizaje automático, que permitan reconocer un conjunto de etiquetas predefinidas según el/los conjunto/s de datos utilizados y medir su desempeño.
- Medir el impacto del uso de la información contextual, presente en el conjunto de datos, en las distintas etapas del proceso de construcción de los clasificadores.

## Aportes de la tesis

El proceso que se presenta en esta tesis, permite agilizar la construcción de clasificadores de emociones en texto basados en aprendizaje automático y a su vez mejorar su desempeño mediante el uso de información contextual. Estos clasificadores pueden ser utilizados para ofrecer una amplia variedad de propósitos potenciales, como detectar la emoción que surge de la opinión de grandes grupos de personas sobre ciertos productos, servicios o incluso políticas públicas. También podrían utilizarse para identificar demandas o quejas no satisfechas de ciudadanos; o, en seguridad, para la detección automática de factores de riesgo en redes sociales, como amenazas, hostigamiento o acoso. Los clasificadores construidos a partir del proceso mencionado, alcanzan un desempeño similar al de otros entrenados con datasets etiquetados manualmente. Debe resaltarse que, en el trabajo presentado, la necesidad de etiquetado manual en el proceso de recolección y clasificación se reduce significativamente.

El conjunto de datos creado puede ser utilizado en diversas investigaciones que realicen Análisis de Sentimientos en español. Además, el proceso de recopilación y validación presentado en esta tesis puede adaptarse fácilmente para generar nuevos datasets en temas o idiomas específicos

## Líneas de I/D futuras

Los siguientes pasos de esta investigación deberían estar enfocados en probar la metodología desarrollada para el Análisis de Sentimientos con:

- Distintas fuentes de datos, lo que permitiría dar certezas acerca de la utilidad de las tareas de preprocesamiento, como así también de los niveles de consenso que pueden lograrse mediante el proceso de validación de etiquetas.
- Diferentes idiomas, debido a que la efectividad de las etiquetas recopiladas puede variar de un idioma al otro, por lo que resulta necesario repetir los experimentos para cada uno de los distintos idiomas. Particularmente sí que persigue el objetivo de realizar Análisis de Sentimientos multilingüaje.
- Variados o múltiples tipos de datos para realizar Análisis de Sentimientos multimodal, lo que requeriría adaptar la metodología desarrollada en esta tesis.
- Diferentes modalidades (presencial / virtual) y longitudes para los cuestionarios de validación de etiquetas, de manera de medir su impacto en los niveles de consenso conseguidos.
- Distintas arquitecturas para los clasificadores, por ejemplo, utilizando las presentadas en la tarea “EmoContext Contextual Emotion Detection in Text” de la competencia SemEval. O también, utilizando otros formatos de representación novedosos para los textos de entrada como el caso de BERT.



**Advisor**  
Dra. Sandra Baldassarri

**Codirector**  
Mg. Hugo Ramon

**Thesis approval date**  
September 27, 2023

**SEDICI**  
[Full Text](#)

## **Modeling and implementation of opinion analysis smart algorithms**

**Keywords:** Sentiment Analysis, Natural Language Processing, Distant Supervision, Machine Learning, Contextual Information

### **Motivation**

Alongside the widespread adoption of social media, the generation of content on these platforms, particularly in text, has also grown. The proliferation of this type of content has provided the necessary raw material to apply text-mining techniques to extract valuable information from the data.

Numerous studies attempting to categorize texts from social media using machine learning classifiers rely on manual content labeling or using pre-labeled public datasets. These approaches have their drawbacks, including the time-consuming process of manually classifying the training data. Another problem is that classifiers are often built using data from different sources than those they analyze. This poses a challenge because if the classifier hasn't been exposed to similar data during the training phase, it will have difficulty categorizing it correctly. Additionally, the availability of resources such as labeled datasets, corpora, or affective dictionaries is limited for languages other than English, restricting the possibilities of constructing aforementioned text classifiers for other languages, including Spanish.

As a result, the collection and validation of resources in the target language become necessary for building supervised machine learning-based text classifiers. However, these tasks are extremely time-consuming and resource-intensive. This problem is exacerbated in cases where the classification criterion is not objective, such as emotion classification in text. In these situations, multiple judges are required to classify the same content to validate the accuracy of the assigned label.

To expedite the development of supervised machine learning-based emotion classifiers for the Spanish language, reducing or eliminating the need for manual labeling of the datasets used for training is necessary. The following specific objectives

are established to achieve that purpose:

- Construct or acquire datasets generated from the interaction on social networks that reflect public opinion on various events, preferably including contextual information about the aforementioned events.
- Carry out cleaning and preprocessing tasks on the input texts to verify their impact on the performance of machine learning-based classifiers.
- Define a methodology and select an appropriate metric to validate the labels assigned to the collected texts.
- Implement machine learning-based text classifiers that recognize a set of predefined labels based on the dataset(s) used, and measure their performance.
- Measure the impact of using contextual information present in the dataset at different stages of the classifier construction process.

### **Thesis contributions**

The process presented in this thesis allows for streamlining the construction of text-based emotion classifiers using machine learning and enhancing their performance using contextual information. These classifiers can be used for a wide variety of potential purposes, such as detecting the sentiment arising from the opinions of large groups of people about specific products, services, or even public policies. They could also be used to identify unmet demands or complaints from citizens or, in security, to automatically detect risk factors in social networks, such as threats, harassment, or bullying.

The classifiers built using the mentioned process perform similarly to others trained with manually labeled datasets. It should be emphasized that in the presented work, the

need for manual labeling in the collection and classification process is significantly reduced.

The constructed dataset can be used for various research purposes involving Sentiment Analysis in Spanish. Furthermore, the collection and validation process presented in this thesis can be easily adapted to generate new resources for specific domains or languages.

### **Future Research Lines**

The following steps of this research should be focused on testing the developed methodology for Sentiment Analysis with the following variations:

- Diverse data sources would provide certainty about the usefulness of preprocessing tasks and the levels of consensus that can be achieved through the label validation process.
- Different languages, as the effectiveness of the collected labels may vary from one language to another, requiring the experiments to be repeated for each language, particularly if the objective is to perform multilingual Sentiment Analysis.
- Diverse or multiple data types to perform multimodal Sentiment Analysis, which would require adapting the methodology developed in this thesis.
- Different modalities (in-person/virtual) and lengths for the label validation questionnaires to measure their impact on the achieved levels of consensus.
- Other architectures for the classifiers, for example, use those presented in the “EmoContext Contextual Emotion Detection in Text” task of the SemEval competition. Alternatively, utilizing other innovative formats for representing the input texts, such as BERT.



**Director**  
Dr. Luis Antonio Olsina

**Codirector**  
Dr. Gustavo Rossi

**Fecha de aprobación**  
30 de octubre de 2023

**SEDICI**  
[Tesis Completa](#)

## **Estrategia Integrada de Pruebas de Software consciente de la Situación y basada en Escenarios**

**Palabras clave:** ontologías; estrategias integradas; pruebas de software; procesos; métodos; situación; contexto; pruebas de escenarios

### **Motivación**

Al llevar a cabo la investigación para este trabajo doctoral, se detectaron oportunidades de mejora relacionadas a estrategias integradas de pruebas de software (o simplemente testing). Notar que una estrategia integrada es aquella que se compone de tres capacidades fundamentales o pilares, a saber: una especificación de proceso; una especificación de método; y una especificación de base conceptual robusta (por ejemplo, estructurada como una ontología). A continuación, se describen las limitaciones observadas que motivaron la realización de este trabajo. Desde lo mejor del conocimiento del autor de esta tesis, actualmente no existen metodologías o estrategias integradas de testing que utilicen una ontología de pruebas de software como base conceptual. Aunque en la literatura se pueden encontrar estrategias como ISO 29119 o ISTQB, que consideran las tres capacidades, estas utilizan un glosario como base conceptual. No obstante, es importante destacar que una ontología es una estructuración semántica más rica que un glosario o taxonomía, ya que no solo define explícitamente los términos principales, sino también sus propiedades, relaciones y axiomas.

Tras realizar una Revisión Sistemática de Literatura (RSL) sobre ontologías de pruebas de software, se concluyó que no existe una ontología de testing adecuada que dé soporte a una familia de estrategias de pruebas de software. Al evaluar las soluciones actuales se detectaron algunas limitaciones como, por ejemplo, no tienen una buena cobertura del dominio considerando términos de pruebas funcionales, no funcionales, estáticas y/o dinámicas; no están directamente relacionadas con conceptos de requisitos no funcionales y/o requisitos funcionales; no tienen calidad estructural satisfactoria; no reutilizan una ontología fundacional; entre otros aspectos.

A pesar de que existen muchos métodos o técnicas para las pruebas de software, no todos ellos son lo suficientemente adecuados para probar el software en diferentes situaciones en las cuales es importante modelar las entidades de contexto que influyen el objeto de prueba. Además, hay un escaso número de tecnologías (es decir, métodos, herramientas, estrategias) que son útiles para probar situaciones en las que es importante considerar las entidades de contexto que interactúan con el objeto de prueba. Por otra parte, a continuación se listan los objetivos planteados para este trabajo:

Objetivo general: Desarrollar una estrategia dirigida por actividades de testing que ayude a las organizaciones a alcanzar propósitos de testing como verificar, validar, encontrar defectos, vulnerabilidades, entre otros. Esta estrategia debe integrar las tres capacidades anteriormente mencionadas, a saber: una especificación de proceso, una especificación de método, y una base conceptual robusta estructurada como una ontología. Además, la estrategia debe ser útil para probar situaciones en las que es importante considerar entidades de contexto. También, la estrategia debe seguir el enfoque de pruebas basadas en escenarios.

Para alcanzar el objetivo general anterior, se propusieron los siguientes objetivos específicos:

Dado que la estrategia a desarrollar debe estar semánticamente soportada por una ontología, se debe investigar cuáles son las ontologías existentes para el dominio de las pruebas de software. Una vez conocido el estado del arte de las ontologías de pruebas de software, se debe adoptar/adaptar alguna existente si cumple con los requisitos establecidos. En caso de que ninguna ontología existente cumpla con los requisitos establecidos, se deberá desarrollar una nueva.

Haciendo uso de los conceptos de pruebas de software de la ontología establecida, se debe desarrollar una especificación de proceso que considere actividades de testing de diseño, ejecución y análisis, y además especifique cuáles son sus artefactos de entrada y salida correspondientes. Usando los conceptos de pruebas de software de la ontología establecida, se debe desarrollar una especificación de método que este basado en la técnica de diseño de pruebas basada en escenarios. Además, debe dar soporte a situaciones que consideren entidades de contexto. Una vez confeccionada la estrategia integrada, se debe validar la misma para inicialmente evaluar su utilidad.

### Aportes de la tesis

A continuación, se listan cada una de las contribuciones realizadas por parte de este trabajo doctoral:

- Una especificación de proceso para revisiones/mapeos sistemáticos de literatura.
- Una RSL sobre ontologías de pruebas de software.
- Una especificación de proceso para la metodología de investigación de Design Science Research.
- Una nueva ontología de dominio de nivel superior para pruebas de software llamada TestTDO (Top-Domain Ontology).
- Una nueva estrategia integrada de testing llamada SaST (Situation-aware Scenario-based Testing Strategy) que fue inicialmente validada por dos estudiantes de grado para sus proyectos finales de ingeniería

### Líneas de I/D futuras

A partir de este trabajo son varias las líneas de investigación en las que se pretende avanzar.

Por un lado, se pretende seguir contribuyendo a la familia de estrategias de testing construyendo nuevas estrategias para diferentes propósitos y que consideren e integren las tres capacidades ya mencionadas. La motivación principal de esta línea de avance futuro es que es muy probable que SaST sea la primera estrategia integrada de pruebas de software que considera una ontología de testing como base conceptual. Por lo tanto, esta familia de estrategias integradas de testing por ahora solo cuenta con una estrategia y podría seguir creciendo incorporando nuevas de ellas.

Relacionado con el trabajo futuro anterior, si se desarrollaran nuevas estrategias integradas de testing más específicas y dado que la ontología TestTDO se encuentra a nivel de dominio superior, se podrían adoptar/adaptar o desarrollar nuevas ontologías de pruebas de software a nivel de dominio inferior para poder cubrir conceptos más específicos. Estas otras ontologías de más bajo nivel se podrían enriquecer con la semántica de dominio brindada por los conceptos de alto nivel de TestTDO. A su vez, se podrían integrar a la arquitectura ontológica llamada

FCD-OntoArch (Foundational, Core, and Domain Ontological Architecture for Sciences), la cual contiene a TestTDO. En este trabajo se realizó una validación inicial de la estrategia SaST al ser aplicada en dos empresas por dos estudiantes de grado para sus proyectos finales de ingeniería. Con la idea de seguir validando y mejorando esta estrategia se pretende aplicarla en otro contexto real para proyectos de software del dominio de las transacciones financieras.



**Advisor**  
Dr. Luis Antonio Olsina

**Codirector**  
Dr. Gustavo Rossi

**Thesis approval date**  
October 30, 2023

**SEDICI**  
[Full Text](#)

## **Situation-aware Scenario-based Integrated Software Testing Strategy**

**Keywords:** ontologies; integrated strategies; software testing; processes; methods; situation; context; scenario-based testing

### **Motivation**

When carrying out the research for this doctoral work, opportunities for improvement related to integrated software testing strategies were detected. Note that an integrated strategy is one that has three fundamental capabilities or pillars, namely: a process specification; a method specification; and a robust conceptual base specification (e.g., structured as an ontology). The observed limitations that led to the realization of this work are described below.

To the best of the knowledge of the author of this thesis, there are currently no integrated testing methodologies or strategies that use a software testing ontology as a conceptual basis. Although strategies such as ISO 29119 or ISTQB can be found in the literature, which considers the three capabilities, they use a glossary as a conceptual basis. However, it is important to note that an ontology is a richer semantic structure than a glossary or taxonomy, since it not only explicitly defines the main terms, but also their properties, relationships and axioms. After conducting a Systematic Literature Review (SLR) on software testing ontologies, it was concluded that there is no adequate testing ontology that supports a family of software testing strategies. When evaluating the current solutions, some limitations were detected, such as, for example, they do not have good coverage of the domain considering terms of functional, nonfunctional, static and/or dynamic testing; they are not directly related to concepts of nonfunctional requirements and/or functional requirements; they do not have satisfactory structural quality; they do not reuse a foundational ontology; in other aspects.

Although there are many methods or techniques for software testing, not all of them are suitable enough to test the software in different situations in which it is important to model the context entities that influence the test object. In

addition, there are a few technologies (i.e., methods, tools, strategies) that are useful for testing situations where it is important to consider the context entities that interact with the test object.

On the other hand, the objectives set for this work are listed below:

General objective: Develop a strategy driven by testing activities that help organizations achieve testing purposes such as verifying, validating, finding defects, vulnerabilities, among others. This strategy must integrate the three aforementioned capabilities, namely: a process specification, a method specification, and a robust conceptual base structured as an ontology. Furthermore, the strategy should be useful for testing situations where it is important to consider context entities. Also, the strategy should follow the scenario-based testing approach.

To achieve the above general objective, the following specific objectives were proposed:

Since the strategy to be developed must be semantically supported by an ontology, it is necessary to research which are the existing ontologies for the software testing domain. Once the state of the art of software testing ontologies is known, an existing one should be adopted/adapted if it meets the established requirements. If no existing ontology meets the established requirements, a new one must be developed. Using the software testing concepts of the established ontology, a process specification must be developed that considers design, execution, and analysis testing activities, and also specifies what their corresponding input and output artifacts are.

Using the software testing concepts of the established ontology, a method specification must be developed based on



the scenario-based test design technique. In addition, it must support situations that consider context entities.

Once the integrated strategy has been built, it must be validated to initially assess its usefulness.

## Thesis contributions

Each of this doctoral work contributions is listed below:

- A process specification for systematic literature mapping/reviews.
- An SLR on software testing ontologies.
- A process specification for Design Science Research methodology.
- A new top-level domain ontology for software testing called TestTDO (Top-Domain Ontology).
- A new integrated testing strategy called SaST (Situation-aware Scenario-based Testing Strategy) which was initially validated by two undergraduate students for their final engineering projects.

## Future Research Lines

From this work, there are several research lines in which it is intended to advance. On the one hand, it is intended to continue contributing to the family of testing strategies by building new strategies for different purposes and that consider and integrate the three capabilities already mentioned. The main motivation for this future work is that SaST is very likely to be the first integrated software testing strategy that considers a testing ontology as its conceptual basis. Therefore, this family of integrated testing strategies currently only has one strategy and could continue to grow by incorporating new ones.

Related to the above future work, if new more specific integrated testing strategies were to be developed and since the TestTDO ontology is at the higher domain level, new software testing ontologies could be adopted/adapted or developed at the lower domain level for to cover more specific concepts. These other lower-level ontologies could be enriched with the domain semantics provided by the higher-level concepts of TestTDO. In turn, they could be integrated into the ontological architecture called FCD-OntoArch (Foundational, Core, and Domain Ontological Architecture for Sciences), which contains TestTDO.

In this work, an initial validation of the SaST strategy was carried out when it was applied in two companies by two undergraduate students for their final engineering projects. With the idea of continuing to validate and improve this strategy, it is intended to apply it in another real context for software projects in the domain of financial transaction.

## DOCTORADO EN CIENCIAS INFORMÁTICAS



**Dr. Martín Becerra**

e-mail

mabecerra@unlam.edu.ar

español

**Director**  
Dr. Jorge Ierache  
Dra. María José Abasolo

**Fecha de aprobación**  
11 de diciembre de 2023

**SEDICI**  
[Tesis Completa](#)

## Framework para la explotación de procedimientos semánticos para navegadores de RA

**Palabras clave:** Realidad Aumentada (RA); Navegadores de RA; Ontologías; Grafos de conocimiento; Web semántica; Procedimientos.

### Motivación

Se plantea como motivación contribuir capacidades faltantes en el estado del arte, una interacción dinámica procedimental para que usuarios puedan emplear contenidos en sus rutinas de la vida cotidiana (Becerra M. et al., Abril 2021).  
Objetivos particulares:

- Proporcionar un framework para navegadores de realidad aumentada semántica que ofrezca la capacidad de interacción procedimental dinámica a usuarios finales para la explotación de procedimientos en contextos de su vida cotidiana.
- Utilizar una ontología como contrato semántico para crear procedimientos con el objetivo de lograr una realidad aumentada universal que unifique sistemas de navegadores de RA heterogéneos.
- Establecer las bases para un diseño de arquitectura del framework propuesto que identifique los principales obstáculos y limitaciones encontradas al aplicar estándares de web semántica a procedimientos para navegadores de RA.
- Discutir los resultados obtenidos del framework propuesto a comparación del resto de las soluciones en el estado del arte.

### Aportes de la tesis

Se busca aportar el diseño de un framework para navegadores de realidad aumentada semántica que le agregue como capacidad faltante del estado del arte una interacción dinámica procedimental a prototipos de navegadores de RA semánticos para que los usuarios finales puedan emplear contenidos en sus rutinas de la vida cotidiana (Becerra M. et al., Abril 2021).

### Líneas de I/D futuras

Integración con herramientas de autor de RA de desarrollo propio. que consiste en extender la arquitectura del presente framework para integrar información a nivel de capa semántica con el Catálogo virtual aumentado y templates(-Mangiarua et. al., 2017)(Becerra, M. et al., 2018)(Mangiarua et. al., 2018).

Línea de investigación de integración de procedimientos en el contexto de gemelos digitales, en la que se propuso agregar una capacidad para extender el framework para poder visualizar información en tiempo real que complemente la explotación de procedimientos



english

**Advisor**  
Dr. Jorge Ierache  
Dra. María José Abasolo

**Thesis approval date**  
December 11, 2023

**SEDICI**  
[Full Text](#)

## **Framework for Interoperable Dynamic Procedure interactions on Semantic Augmented Reality Browsers**

**keywords:** Augmented Reality (AR); AR Browsers; Ontology; Knowledge graphs; Semantic Web; Procedures

### **Motivation**

In this Phd thesis is proposed as motivation to contribute missing capacities in the state of the art, a procedural dynamic interaction so that users can use content in their daily life routines (Becerra M. et al., April 2021).

Particular objectives:

- Provide a framework for semantic augmented reality browsers that offers the capability of dynamic procedural interaction to end users for the exploitation of procedures in contexts of their daily life.
- Using an ontology as a semantic contract to create procedures with the aim of achieving a universal augmented reality that unifies heterogeneous AR browser systems.
- Establish the bases for an architectural design of the proposed framework that identifies the main obstacles and limitations encountered when applying semantic web standards to procedures for AR browsers.
- Discuss the results obtained from the proposed framework compared to the rest of the solutions in the state of the art.

### **Thesis contributions**

This Phd thesis aims to contribute the design of a framework for semantic augmented reality browsers that adds dynamic procedural interaction to prototypes of semantic AR browsers as a missing state-of-the-art capability so that end users can use content in their daily life routines. (Becerra M. et al., April 2021).

### **Future Research Lines**

Integration with self-developed AR authoring tools. which consists of extending the architecture of this framework to integrate information at the semantic layer level with the augmented Virtual Catalog and templates (Mangiarua et. al., 2017) (Becerra, M. et al., 2018)(Mangiarua et. al. , 2018).

Line of investigation of integration of procedures in the context of digital twins, in which it was proposed to add a capacity to extend the framework to be able to visualize information in real time that complements the exploitation of procedures.



02

MAESTRÍAS

INTELIGENCIA DE DATOS  
ORIENTADA A BIG DATA

CÓMPUTO DE ALTAS  
PRESENTACIONES

INGENIERÍA DE SOFTWARE

REDES DE DATOS

TECNOLOGÍA INFORMÁTICA  
APLICADA EN EDUCACIÓN



**Director**  
Dr. Marcelo Errecalde

**Codirector**  
Mg. Mario Oloriz

**Fecha de aprobación**  
14 de abril de 2023

**SEDICI**  
[Tesis Completa](#)

## **Clasificación automática de correos electrónicos**

**Palabras clave:** Clasificación automática; aprendizaje automático; clasificación semi-supervisada; datos desbalanceados multiclase

### **Motivación**

El objetivo general de este trabajo consiste en estudiar y analizar el conocimiento existente sobre técnicas aprendizaje automático aplicadas a la clasificación automática de textos, particularmente de correos electrónicos, y generar un modelo que aborde un problema concreto. Esto trae aparejados los siguientes objetivos específicos: - Analizar, describir y sistematizar el estado del arte de la clasificación automática de correos electrónicos.

- Diseñar un proceso general para el tratamiento y clasificación automática de correos electrónicos, intentando categorizar esta problemática dentro de la disciplina general de Minería de Textos, la cual abarca las características y particularidades que se originan en esta forma de comunicación.
- Abordar un estudio experimental a partir del procesamiento de las consultas que los estudiantes de la Universidad Nacional de Luján formulan, mediante correo electrónico, asegurando la calidad de los datos y etiquetando, a partir de especialistas en el dominio, un subconjunto de los mismos con las temáticas a las que corresponden en ese dominio.
- Consolidar, a partir de lo anterior, una base de conocimiento con los correos electrónicos etiquetados, representados a partir de diferentes estrategias de representación de textos incorporando a los mismos un proceso de curado y generación de atributos estáticos.
- Entrenar un modelo para la clasificación automática de estos correos electrónicos, abordando de manera concreta esta problemática y dimensionando el problema.
- Indagar en el estado del arte de la clasificación se-

mi-supervisada de documentos, proponiendo una estrategia de esta índole para mejorar la performance de clasificación del modelo entrenado.

### **Aportes de la tesis**

En este trabajo se aborda el desafío de analizar, describir y sistematizar el estado del arte del email mining, puntualmente el área que aborda la clasificación automática de correos electrónicos.

Además, se abordó la construcción de un clasificador automático sobre un caso de estudio real que incluyó 1000 correos de entrenamiento divididos en 16 clases. Este conjunto de datos está conformado con las consultas que realizan los estudiantes al staff administrativo de la Universidad Nacional de Luján sobre temas de índole académica. Se procesó y curó una base con 20.000 correos electrónicos y sus respectivas respuestas, asegurando la calidad de los datos y etiquetando, a partir de especialistas en el dominio, un subconjunto de 1000 correos con las temáticas que abordan.

En este sentido, se entrenó un modelo para la clasificación automática de estos correos electrónicos, abordando de manera concreta esta problemática, alcanzando un accuracy del 86% en el mejor de los casos.

Además, con el objetivo de buscar alternativas para mejorar los modelos entrenados, se indagó en el estado del arte de la clasificación semi-supervisada de documentos, logrando, como una de las principales contribuciones de este trabajo, proponer una nueva estrategia para mejorar la performance de clasificación de los modelos

entrenados La hipótesis acuñada para la definición de esta estrategia es que la selección de características representativas de cada clase, combinada con enfoques de recuperación de información, constituyen un método semisupervisado válido y sencillo para la clasificación automática de correos electrónicos. Desde allí, se generó un proceso para la clasificación semi-supervisada de correos electrónicos a partir de la identificación de características claves de cada clase, utilizando tres técnicas de selección de características como la regresión logística, TF-IDF y SS3 y la posterior recuperación de correos etiquetados automáticamente a partir de un enfoque de recuperación de información con un motor de búsqueda de propósito general como Elasticsearch. Además, se pudo verificar que las estrategias propuestas para el etiquetado semi-supervisado permiten identificar los términos más representativos de cada correo electrónico, al mismo tiempo que es posible utilizar las ponderaciones definidas por cada técnica para valorar la representatividad de esos términos para cada clase. Asimismo, se demostró para el conjunto de datos con que se trabajó que estas técnicas de selección de características, utilizadas como estrategias de etiquetado automático a partir de un enfoque de recuperación de información para la clasificación semi-supervisada, mejoran la capacidad de los clasificadores cuando se incorporan las instancias etiquetadas automáticamente a las etiquetadas de forma manual para entrenar el modelo, alcanzando mejoras de entre el 2% y el 6% para muchas de las estrategias abordadas en el marco de esta investigación.

Por último, se presentó, como apéndice, nuevamente la estrategia como una estrategia de oversampling para aplicar al aprendizaje automático en ambientes de datos desbalanceados.

En este sentido, nuevamente se demostró, al menos para los datos con que se evaluó, que esta nueva estrategia presentada es competitiva en relación a las estrategias de remuestreo para el balanceo de clases, tanto de undersampling como de oversampling, arrojando valores más altos para las métricas de selección de modelos utilizadas y el conjunto de datos sobre el que se realizaron los experimentos.

### Líneas de I/D futuras

Existen diversas líneas de investigación que quedan abiertas a partir de los resultados de este trabajo:

- Los correos electrónicos en general, y este conjunto de datos en particular, tienen características que no

benefician a los modelos de clasificación automática dada la informalidad y los errores de sintaxis recurrentes propios de la dinámica de este medio de comunicación, por lo cual se espera que en conjuntos de datos con textos más depurados se obtengan modelos que logren mejores resultados, para lo cual sería interesante incorporar un proceso de corrección ortográfica sobre el corpus de documentos para la depuración del lenguaje.

- En este trabajo se propone un enfoque semi-supervisado para el etiquetado y clasificación de documentos y se demuestra que el mismo es un enfoque válido y sencillo, al menos para esta colección de documentos. Queda como desafío futuro abordar ajustes en este proceso sobre los parámetros involucrados, ya sea en términos de las técnicas de selección de características como el ajuste fino del  $\$N\$$  con la cantidad de términos representativos para cada clase y el impacto de su calibración en la eficiencia y eficacia de esta estrategia.

- A su vez, para la consolidación de la estrategia de aprendizaje semi-supervisado propuesta, quedan pendientes nuevas pruebas empíricas en otros conjuntos de datos que permitan ratificar la utilidad de la misma, al mismo tiempo que ayude a identificar los contextos más propicios para su utilización.

- Asimismo, en este trabajo se propone enfoque anterior también como estrategia de oversampling para la clasificación automática en un ambiente de datos con clases desbalanceadas. Si bien en el presente estudio se ha limitado la experimentación al dominio de la clasificación automática de correos electrónicos, creemos que la estrategia propuesta es generalizable a otros dominios donde existan disponibles documentos de textos sin etiquetar y, como trabajo futuro, se propone realizar un nuevo trabajo aplicado a un contexto más general de clasificación automática de textos.

- Por otro lado, se plantea la posibilidad de avanzar en la implementación de un modelo de clasificación automática en un entorno de aplicación real, con el fin de realizar una transferencia concreta para facilitar gestión de las respuestas a las consultas en el contexto de la Universidad Nacional de Luján, así como también, analizar la posibilidad de adaptar esta solución -al menos en términos del proceso de construcción propuesto- a otras problemáticas de índole similar.



**Advisor**  
Dr. Marcelo Errecalde

**Codirector**  
Mg. Mario Oloriz

**Thesis approval date**  
April 14, 2023

**SEDICI**  
[Full Text](#)

## Automatic email classification

**Keywords:** Text classification; machine learning; semi-supervised text classification; multi-class unbalanced data

### Motivation

The general objective of this work is to study and analyze the existing knowledge on machine learning techniques applied to automatic classification of texts, particularly e-mails, and to generate a model that addresses a specific problem. This entails the following specific objectives:

- Analyze, describe and systematize the state of the art of automatic classification of emails.
- Design a general process for the automatic treatment and classification of emails, trying to categorize this problem within the general discipline of Text Mining, which covers the characteristics and particularities that originate in this form of communication.
- Address an experimental study based on the processing of the queries that the students of the National University of Luján formulate, through email, ensuring the quality of the data and labeling, from specialists in the domain, a subset of them with the themes to which they correspond in that domain.
- Consolidate, from the above, a knowledge base with labeled e-mails, represented from different text representation strategies, incorporating a curing process and generation of static attributes.
- Train a model for the automatic classification of these emails, specifically addressing this problem and sizing the problem.
- Investigate the state of the art of semi-supervised document classification, proposing a strategy of this nature to improve the classification performance of the trained model.

### Thesis contributions

This paper addresses the challenge of analyzing, describing and systematizing the state of the art of email mining, specifically the area that deals with the automatic classification of emails. In addition, the construction of an automatic classifier on a real case study that included 1000 training emails divided into 16 classes was addressed. This dataset is made up of queries made by students to the administrative staff of the National University of Lujan on academic topics. A database with 20,000 emails and their respective responses was processed and curated, ensuring the quality of the data and labeling, based on domain specialists, a subset of 1000 emails with the topics they address.

In this sense, a model for the automatic classification of these e-mails was trained, addressing this problem in a concrete way, reaching an accuracy of 86% in the best case. In addition, with the aim of seeking alternatives to improve the trained models, the state of the art of semi-supervised document classification was investigated, achieving, as one of the main contributions of this work, to propose a new strategy to improve the classification performance of the trained models. The hypothesis coined for the definition of this strategy is that the selection of representative features of each class, combined with information retrieval approaches, constitute a valid and simple semi-supervised method for the automatic classification of e-mails. From there, a process for semi-supervised email classification was generated from the identification of key features of each class, using three feature selection



techniques such as logistic regression, TF-IDF and SS3 and the subsequent retrieval of automatically labeled emails from an information retrieval approach with a general purpose search engine such as Elasticsearch.

In addition, it was possible to verify that the strategies proposed for semi-supervised tagging allow the identification of the most representative terms of each email, while it is possible to use the weights defined by each technique to assess the representativeness of these terms for each class.

It was also demonstrated for the dataset worked with that these feature selection techniques, used as automatic labeling strategies from an information retrieval approach for semisupervised classification, improve the capacity of the classifiers when the automatically labeled instances are incorporated to the manually labeled ones to train the model, reaching improvements between 2% and 6% for many of the strategies addressed in the framework of this research.

Finally, the strategy was presented, as an appendix, again as an oversampling strategy to apply to machine learning in unbalanced data environments. In this sense, it was again demonstrated, at least for the data with which it was evaluated, that this new strategy presented is competitive in relation to the resampling strategies for class balancing, both undersampling and oversampling, yielding higher values for the model selection metrics used and the data set on which the experiments were performed.

### Future Research Lines

There are several lines of research that remain open from the results of this work:

- Emails in general, and this dataset in particular, have characteristics that do not benefit automatic classification models given the informality and recurrent syntax errors typical of the dynamics of this communication medium, so it is expected that in datasets with more refined texts models that achieve better results are obtained, for which it would be interesting to incorporate a process of spelling correction on the corpus of documents for language debugging.
- This paper proposes a semi-supervised approach for the labeling and classification of documents and demonstrates that it is a valid and simple approach, at least for this collection of documents. It remains as a future challenge to address adjustments in this process

on the parameters involved, either in terms of the feature selection techniques such as fine tuning the N with the number of representative terms for each class and the impact of its calibration on the efficiency and effectiveness of this strategy.

- In turn, for the consolidation of the proposed semi-supervised learning strategy, further empirical tests on other data sets are pending to ratify its usefulness, while helping to identify the most favorable contexts for its use.
- Likewise, in this work, the above approach is also proposed as an oversampling strategy for automatic classification in a data environment with unbalanced classes. Although in the present study the experimentation has been limited to the domain of automatic classification of e-mails, we believe that the proposed strategy is generalizable to other domains where unlabeled text documents are available and, as a future work, it is proposed to carry out a new work applied to a more general context of automatic classification of texts.
- On the other hand, the possibility of advancing in the implementation of an automatic classification model in a real application environment is proposed, in order to make a concrete transfer to facilitate the management of responses to queries in the context of the Universidad Nacional de Luján, as well as to analyze the possibility of adapting this solution -at least in terms of the proposed construction process- to other problems of a similar nature.



**Directores**

Dr. Francisco Igual  
Dra. Laura De Giusti

**Fecha de aprobación**  
4 de mayo de 2023

SEDICI  
[Tesis Completa](#)

**Coplanificación de procesos  
maleables de aprendizaje  
automático mediante  
contenedores**

**Palabras clave:** Coplanificación; aplicaciones elásticas; tensorflow; contenedores; HPC; procesos elásticos

## Motivación

El crecimiento exponencial en interés por el aprendizaje automático en la última década está directamente relacionado con tres avances fundamentales:

1. El desarrollo de mejores algoritmos con aplicaciones directas en muchos campos de la ciencia y la ingeniería;
2. La disponibilidad de cantidades masivas de datos y la viabilidad de almacenarlos y analizarlos de manera eficiente;
3. La aparición de arquitecturas de hardware novedosas, normalmente paralelas y / u homogéneas, que permiten una adecuada explotación de estos algoritmos sobre grandes conjuntos de datos en un tiempo asequible.

La aparición de nuevas arquitecturas y técnicas en HPC ha renovado el interés por el Machine Learning en una gran variedad de problemas, incluyendo aplicaciones de reconocimiento de imágenes, segmentación, reconocimiento de voz, procesamiento de lenguaje natural o traducción de idiomas, entre muchos otros. Tensorflow, Caffe, Keras y PyTorch son frameworks de ML que permiten ocultar detalles de implementación al usuario manteniendo un alto rendimiento tanto en el entrenamiento como en la inferencia de modelos.

Hoy en día, Tensorflow es uno de los frameworks más utilizados. Su diseño se basa en un grafo de ejecución en el cual las operaciones están representadas por nodos y el flujo de datos por tensores (matrices multidimensionales de datos). Para cada operación puede definirse el paralelismo (intra paralelismo). A su vez, se puede elegir la cantidad de operaciones simultáneas que pueden ejecutarse (inter paralelismo). El gran problema que presentan ambos paralelismos es que son

definidos antes de la ejecución del algoritmo sin posibilidad de modificarlos en tiempo de ejecución.

La rigidez en el uso de recursos no es en sí misma problemática, siempre que las aplicaciones se ejecuten de forma aislada en el computador. Sin embargo, en entornos en los que múltiples aplicaciones conviven en el mismo computador, y tanto sus puntos de llegada al sistema como los requisitos a nivel de recurso que éstas presentan puedan variar, dicha rigidez puede llevar a situaciones de sobre suscripción, degradando el rendimiento global del sistema y particular de cada aplicación.

Así, parece necesario no solo aplicar técnicas de planificación de procesos y gestión de recursos avanzadas que permitan optimizar el rendimiento global del sistema y particular de las aplicaciones, respetando a la vez los requisitos de calidad de servicio (QoS) sino además hacerlo siendo consciente de la posibilidad de gestionar aplicaciones elásticas.

Por otro lado, la utilización de contenedores como método de virtualización ligera permite una mejor administración de los recursos y portabilidad. Existen múltiples planificadores que permiten aprovechar los beneficios de los contenedores, pero solo permiten llevar a cabo una asignación estática de recursos en el momento de su creación, y en algunos casos reasignación de recursos en tiempo de ejecución; en cualquier caso, las aplicaciones en ejecución dentro del contenedor no se encuentran preparadas para reaccionar ante dicho evento, y por tanto no se adaptarán en ningún caso a la modificación en los recursos asignados al contenedor.

Los objetivos de este trabajo son:

- Modificación del esquema de gestión de recursos dentro de la infraestructura del framework Tensorflow para permitir selección dinámica del paralelismo de las operaciones que conforman el modelo de ML.
- Diseño e implementación de un controlador interno de cada contenedor que permita manejar los recursos computacionales asignados dinámicamente al algoritmo de ML, y de un mecanismo de comunicación entre el sistema y la aplicación confinada en el contenedor.
- Diseño e implementación de un planificador de contenedores que ejecutan algoritmos de ML sobre Tensorflow elástico utilizando técnicas de orquestación que permitan administrar los recursos computacionales del sistema eficientemente.

### **Aportes de la tesis**

En este trabajo se propone el diseño e implementación de un mecanismo completo de elasticidad en el uso de recursos computacionales en el framework Tensorflow, permitiendo la reasignación dinámica de núcleos de cómputo durante la ejecución del algoritmo de ML. Además, se extiende el uso de la elasticidad a contenedores con la implementación de un controlador/cliente que permita administrar los recursos computacionales asignados a los algoritmos de ML que se ejecutan internamente. Por último, se implementa un planificador de contenedores elásticos con el fin de gestionar dinámicamente los recursos del sistema entre todos los contenedores activos y definir políticas de planificación que favorezcan el rendimiento global del sistema.

### **Líneas de I/D futuras**

- Incorporación de nuevas políticas de planificación de contenedores que permitan mejorar el rendimiento alcanzado para las implementadas anteriormente.
- Estudio y análisis de la eficiencia energética del coplanificador.
- Incorporar en el planificador soporte de sistemas heterogéneos utilizando aceleradores tales como GPU, TPU, entre otros.
- Analizar el uso de funciones desarrolladas por proveedores de GPU, como Multi-Stream, Multi-Process Service (MPS), Multi-Instance GPU (MIG) y GPUs virtuales (vCS) para soporte de planificación en tiempo de ejecución y administración de recursos.
- Explorar la administración de recursos en otros frameworks de ML tales como PyTorch y Caffe, con el fin de extender la elasticidad a más aplicaciones y realizar estudios comparativos entre ellos.



**Advisors**

Dr. Francisco Igual  
Dra. Laura De Giusti

**Thesis approval date**

May 4, 2023

**SEDICI**

[Full Text](#)

## **Scheduling elastic machine learning process through containers**

**Keywords:** Coscheduling; elastic applications; tensorflow; containers; HPC; elastic processes

### **Motivation**

The exponential growth in interest in machine learning over the last decade is directly related to three fundamental advances:

1. The development of better algorithms with direct applications in many fields of science and engineering;
2. The availability of massive amounts of data and the feasibility of storing and analyzing it efficiently;
3. The emergence of novel hardware architectures, usually parallel and/or homogeneous, that allow adequate exploitation of these algorithms on large data sets in an affordable time.

The emergence of new architectures and techniques in HPC has renewed interest in Machine Learning for a wide variety of problems, including applications in image recognition, segmentation, speech recognition, natural language processing or language translation, among many others. Tensorflow, Caffe, Keras and PyTorch are ML frameworks that allow hiding implementation details from the user while maintaining high performance in both training and model inference.

Nowadays, Tensorflow is one of the most widely used frameworks. Its design is based on an execution graph in which the operations are represented by nodes and the data flow by tensors (multidimensional data matrices). Parallelism (intra-parallelism) can be defined for each operation. Also, the inter parallelism allows for the selection of the desired number of simultaneous operations that can be executed. The big problem with both parallelisms is that they are defined before the execution of the algorithm without the possibility of modifying them at runtime.

Static resource usage is not in itself problematic, as long as applications run in isolation on the computer. However, in environments where multiple applications coexist on the same computer, and both their arrival points to the system and their resource requirements may vary, such rigidity can lead to oversubscription situations, degrading the overall system performance and the performance of individual applications.

Thus, it seems necessary not only to apply advanced scheduling process and resource management techniques to optimize overall system and application performance, while respecting quality of service (QoS) requirements, but also to do so while being aware of the possibility of managing elastic applications.

However, the use of containers as a lightweight virtualization method allows for better resource management and portability. There are multiple schedulers that allow taking advantage of the benefits of containers, but they only allow carrying out a static allocation of resources at the time of their creation, and in some cases reallocation of resources at resource time.

In any case, the applications running inside the container are not prepared to react to such an event, and therefore will not adapt in any case to the modification in the resources assigned to the container.

The objectives of this work are:

- Modification of the resource management scheme within the Tensorflow framework infrastructure to allow dynamic selection of the parallelism of the operations that make up the ML model.

- Design and implementation of an internal controller for each container to manage the computational resources dynamically assigned to the ML algorithm, and of a communication mechanism between the system and the application confined in the container.
- Design and implementation of a scheduler for containers running ML algorithms on elastic Tensorflow using orchestration techniques to efficiently manage the computational resources of the system.

### Thesis contributions

This paper proposes the design and implementation of a complete mechanism of elasticity in the use of computational resources in the Tensorflow framework, allowing the dynamic reallocation of computational cores during the execution of the ML algorithm. Furthermore, the use of elasticity is extended to containers with the implementation of a controller/client to manage the computational resources allocated to ML algorithms running internally. Finally, an elastic container scheduler is implemented in order to dynamically manage system resources among all active containers and define scheduling policies that favor overall system performance.

### Future Research Lines

- Incorporation of new container scheduling policies to improve the performance achieved for those previously implemented.
- Study and analysis of the energy efficiency of the co-planner.
- Incorporate in the scheduler support for heterogeneous systems using accelerators such as GPU, TPU, among others.
- Analyze the use of features developed by GPU vendors, such as Multi-Stream, Multi-Process Service (MPS), Multi-Instance GPU (MIG) and virtual GPUs (vCS) for runtime scheduling support and resource management.
- Explore resource management in other ML frameworks such as PyTorch and Caffe, in order to extend elasticity to more applications and perform comparative studies between them.



**Director**  
Dr. Andrés Rodríguez

**Codirectora**  
Dra. Alejandra Garrido

**Asesor Científico**  
Mg. Matías Butti

**Fecha de aprobación**  
30 de mayo de 2023

**SEDICI**  
[Tesis Completa](#)

## **Soporte al proceso de análisis e interpretación de variantes genéticas en Exoma humano**

**Palabras clave:** Ingeniera en Sistemas de Información; Desarrolladora; Analista Funcional; Posgrado de Inteligencia de Datos orientada a Big Data

### **Motivación**

El propósito principal de esta tesis es dar soporte al médico genetista en el proceso de diagnóstico genético, en particular el acceso a la información sobre la composición genética de sus pacientes, la identificación de los posibles escenarios de detección y diagnóstico y el registro de cada uno de los pasos y datos llevados a cabo en el proceso.

#### **Objetivos específicos**

- Modelar el proceso de toma de decisiones para el filtrado y selección de variantes de un exoma relevantes llevado a cabo por el médico genetista para alcanzar un informe de diagnóstico de un paciente.
- Proveer herramientas de asistencia, basadas en una visualización eficaz, para la ejecución de ese proceso de análisis e interpretación con mayor rapidez y sensibilidad.
- Acelerar los tiempos de interpretación.
- Aumentar la sensibilidad (en este caso la tasa de detección de las variantes patogénicas causales de la enfermedad o que sean relevantes para su salud).

### **Aportes de la tesis**

Esta tesis aporta a la comunidad médica, dedicada al área genética, elementos para poder realizar un diagnóstico integral en base al análisis del Exoma Humano y generar una base de datos genéticos propia con el registro del trabajo realizado, pudiendo de esta manera

mejorar los tiempos de evaluación diagnóstica, tener la información centralizada y contar con información para diagnósticos futuros.

### **Líneas de I/D futuras**

Como principal trabajo futuro se sugiere que, en caso de ser necesario porque el volumen de trabajo lo justifique, sea integrada la consulta manual a las bases de datos de datos genéticas tales como OMIM, Phenomizer, entre otras. Podría también construirse un sistema de recomendación basado en la lectura previa de las mencionadas apis.

Podría también a futuro ser desarrollado un módulo de contribución e integración a las bases de datos de estadísticas de enfermedades detectadas en la población argentina, a efectos de tomar decisiones de prevención en las patologías más frecuentes. El diseño de la interfaz de usuario puede utilizarse como punto de partida para un proceso de desarrollo e implementación de la herramienta Exoma Browser.



**Advisor**

Dr. Andrés Rodríguez

**Codirector**

Dra. Alejandra Garrido

**Scientist Consultant**

Mg. Matías Butti

**Thesis approval date**

May 30, 2023

**SEDICI**

[Full Text](#)

**Support to the process of analysis  
and interpretation of genetic  
variants in human Exome**

**Keywords:** Information Systems Engineer; Developer; Functional Analyst; Postgraduate in Data Intelligence oriented to Big Data

**Motivation**

The main purpose of this thesis is to support the geneticist doctor in the process of genetic diagnosis, in particular the access to information about the genetic composition of their patients, the identification of possible scenarios of detection and diagnosis and the recording of each of the steps and data carried out in the process.

**Specific objectives**

- Model the decision-making process for the filtering and selection of relevant exome variants carried out by the geneticist doctor to reach a diagnostic report for a patient.
- Provide assistance tools, based on efficient visualization, for executing the analysis and interpretation process more quickly and sensitively.
- Accelerate interpretation times.
- Increase sensitivity (in this case the detection rate of pathogenic variants that are causative of the disease or relevant to its health).

**Thesis contributions**

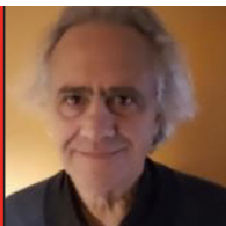
This thesis provides to the medical community, dedicated to the genetic area, with elements to be able to perform an integral diagnosis based on the analysis of the Human Exome and to generate its own genetic database with the record of the work performed, thus being able to improve diagnostic evaluation times, to

have centralized information and to have information for future diagnoses.

**Future Research Lines**

As the main future work, it is suggested that, if it is necessary, because the volume of work justifies it, manual consultation of genetic databases such as OMIM, Phenomizer, among others, could be integrated. A recommendation system could also be built based on the previous reading of the mentioned apis.

In the future, a module for contribution and integration to the statistical databases of diseases detected in the Argentine population could also be developed, in order to make preventive decisions on the most frequent pathologies. The design of the user interface can be used as a starting point for a process of development and implementation of the Exoma Browser tool.



español

**Directora**  
Dra. Massa Stella Maris

**Codirector**  
Dr. Gustavo Hector Rossi

**Fecha de aprobación**  
18 de agosto de 2023

**SEDICI**  
[Tesis Completa](#)

## Elicitación de requerimientos centrada en el usuario para el desarrollo de Serious Games

**Palabras clave:** Elicitación; Serious Games; Educación; Modelos de Proceso

### Motivación

El uso de videojuegos como herramienta educativa, es anterior al comienzo de este siglo. Sin embargo no ha logrado imponerse, a pesar de los esfuerzos puestos en ello. Las mejores propuestas carecen de la inmersión y jugabilidad necesarias. Y esta carencia no les permite trascender el aula a través de su valor lúdico, potenciando su potencial de aprendizaje. En opinión del grupo GTI esto sucederá de la mano de la industria de VJ, por lo cual es preciso dotarla con las herramientas adecuadas. Por eso propone un modelo para el desarrollo de SG en el aula y dentro de este un mecanismo de elicitación, adecuado a las características colaborativas y multidisciplinarias del producto.

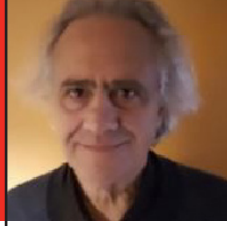
### Aportes de la tesis

- Se ha identificado el tipo de requerimiento específico de un serious game (la mecánica SGM) y sus componentes.
- En función de la naturaleza multidisciplinaria y colaborativa de los videojuegos, se ha diseñado un mecanismo de elicitación. El cual potencia la colaboración y el intercambio interdisciplinario de ideas y puntos de vista.

### Líneas de I/D futuras

- Realizar estudios interdisciplinarios para aumentar la eficacia y precisión en la elicitación de las SGM (mecánicas de juego de los serious game).
- Realizar estudios para una mayor adecuación de las técnicas de elicitación propuestas al producto específico, incorporando los contenidos multimedia, tales como efectos y sonido.
- Profundizar en las técnicas de elicitación, respecto a las analíticas de aprendizaje y el mapeo LM-GM.
- Especificar, diseñar e implementar una herramienta que permita, la implementación y trazabilidad del mecanismo de elicitación.





english

**Advisor**  
Dra. Massa Stella Maris

**Codirector**  
Dr. Gustavo Hector Rossi

**Thesis approval date**  
August 18, 2023

**SEDICI**  
[Full Text](#)

## **User-based requirements elicitation to develop serious games**

**Keywords:** Elicitation; Serious Games; Education; Process Models

### **Motivation**

The use of video games as an educational tool is prior to the beginning of this century. However, it has not managed to prevail, despite the efforts put into it. The best proposals lack the necessary immersion and gameplay. And this lack does not allow them to transcend the classroom through their playful value, enhancing their learning potential. It is the opinion of the GTI group that this will happen hand in hand with the VJ industry, which is why it is necessary to provide it with the appropriate tools. For this reason, it proposes a model for the development of SG in the classroom and within this an elicitation mechanism, appropriate to the collaborative and multidisciplinary characteristics of the product.

### **Thesis contributions**

- The type of specific requirement of a serious game (SGM mechanics) and its components have been identified.
- Based on the multidisciplinary and collaborative nature of video games, an elicitation mechanism has been designed. Which promotes collaboration and the interdisciplinary exchange of ideas and points of view.

### **Future Research Lines**

- Interdisciplinary studies for to increase the efficiency and precision in eliciting SGM (serious game mechanics).
- Studies to better adapt the elicitation techniques proposed, incorporating multimedia content, such as effects and sound.
- Improve the integration of learning analytics and LM-GM mapping.
- Specify, design and implement a tool that assists in the implementation and traceability of the elicitation mechanism.



**Directores**

Dra. Laura Lanzarini  
Dr. Franco Ronchetti

**Fecha de aprobación**  
15 de noviembre de 2023

**SEDICI**  
[Tesis Completa](#)

## Minería de Procesos en la Ingeniería de Software Análisis del proceso de distribución de piezas postales

**Palabras clave:** Minería de Procesos; Minería de Datos; Distribución Postal; Proceso Postal; Gestión de Procesos de Negocio

### Motivación

Durante mi desempeño de más de 12 años en desarrollo de Sistemas de Información para el sector postal tuve la oportunidad de conocer detalladamente el proceso de entrega de varios productos con diferentes características. Al no encontrar una herramienta capaz de solucionar o disminuir el impacto de los inconvenientes de distribución decidí realizar un análisis aplicando Ingeniería de Software y Minería de Procesos para generar una solución.

El objetivo central de esta tesis es aplicar las técnicas de Minería de Procesos en un contexto específico como es la distribución postal en la República Argentina utilizando una herramienta desarrollada con este fin que permita encontrar desvíos operativos y cuellos de botella de forma automatizada.

### Aportes de la tesis

Los resultados obtenidos demostraron es posible aplicar las técnicas de Minería de Procesos en un contexto específico como es la distribución postal en la República Argentina de forma manual o a través de una herramienta informática que permita incorporar este análisis al proceso operativo.

Una vez modelizado el proceso e identificados los puntos críticos, como parte de esta investigación, se desarrolló un prototipo de software completamente funcional que permite la aplicación de la Minería de Procesos en el circuito productivo actual. Este prototipo ofrece a los actores involucrados en la distribución postal una herramienta poderosa para analizar y monitorear el rendimiento del proceso, identificar oportunidades de mejora y tomar decisiones basadas en datos con el objetivo de maximizar la eficiencia operativa.

Se generó una herramienta que automatiza y facilita el análisis de los envíos que están en proceso permitiendo una rápida detección de incidencias operativas y facilitando su resolución.

### Líneas de I/D futuras

Implementación de técnicas de aprendizaje automático (machine learning) en conjunto con la Minería de Procesos en el contexto de la distribución postal. Esto permitiría aprovechar el potencial de los algoritmos de aprendizaje automático para predecir desvíos operativos de forma automatizada, identificar patrones ocultos y ofrecer recomendaciones para optimizar el proceso de distribución.

Incorporación de tecnologías emergentes como Internet de las cosas (IoT) y la inteligencia artificial (IA) en la distribución postal. Por ejemplo, el uso de sensores en los vehículos de entrega para recopilar datos en tiempo real y la aplicación de algoritmos de IA para la toma de decisiones automatizada en situaciones complejas.

Otra línea de investigación interesante sería explorar enfoques de optimización y simulación en la distribución postal. Estos enfoques permitirían modelar y simular diferentes escenarios operativos, teniendo en cuenta variables como la carga de trabajo, la disponibilidad de recursos y las restricciones logísticas. Esto ayudaría a identificar configuraciones óptimas y estrategias de planificación que maximicen la eficiencia y la satisfacción del cliente.



**Advisors**  
Dra. Laura Lanzarini  
Dr. Franco Ronchetti

**Thesis approval date**  
November 15, 2023

**SEDICI**  
[Full Text](#)

## **User-centered requirements elicitation for the development of Serious Games**

**Keywords:** Process Mining; Postal distribution ; Process postal; Data mining; Business Process Management

### **Motivation**

For over 12 years, I worked in the development of Information Systems for the postal sector, where I had the opportunity to thoroughly understand the delivery process of various products. Unable to find a tool capable of solving or reducing the impact of distribution issues, I decided to conduct analysis using Software Engineering and Process Mining to generate a solution.

The main objective of this thesis is to apply Process Mining techniques in a specific context, such as postal distribution in the Argentine Republic, using a tool developed for this purpose. This tool aims to automatically identify operational deviations and bottlenecks.

### **Thesis contributions**

The obtained results demonstrated that it is possible to apply Process Mining techniques in a specific context, such as postal distribution in the Argentine Republic, either manually or through a computer tool that allows incorporating this analysis into the operational process.

Once the process was modeled and critical points identified, as part of this research, a fully functional software prototype was developed that enables the application of Process Mining in the current production circuit. This prototype provides stakeholders in postal distribution with a powerful tool to analyze and monitor process performance, identify improvement opportunities, and make data-driven decisions to maximize operational efficiency.

An automated tool was created that facilitates the analysis of shipments in progress, allowing for quick detection of operational incidents and facilitating their resolution.

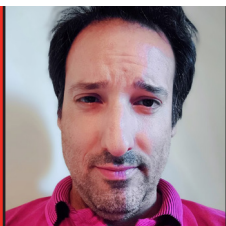
### **Future Research Lines**

Implementation of machine learning techniques in conjunction with Process Mining in the context of postal distribution. This would harness the potential of machine learning algorithms to predict operational deviations automatically, identify hidden patterns, and provide recommendations to optimize the distribution process.

Incorporation of emerging technologies such as the Internet of Things (IoT) and artificial intelligence (AI) in postal distribution. For example, the use of sensors in delivery vehicles to collect real-time data and the application of AI algorithms for automated decision-making in complex situations.

Another interesting research avenue would be to explore optimization and simulation approaches in postal distribution. These approaches would allow modeling and simulating different operational scenarios, considering variables such as workload, resource availability, and logistical constraints. This would help identify optimal configurations and planning strategies that maximize efficiency and customer satisfaction.

## MAESTRÍA REDES DE DATOS



**Mg. Marcelo Alberto Gómez**

e-mail

mgomez@unsj.edu.ar

español

**Directora**  
Dra. Patricia Bazán

**Codirector**  
Mg. Nicolás del Río

**Asesor Científico**  
Ing. Miguel Ángel Morandi

**Fecha de aprobación**  
17 de octubre 2023

**SEDICI**  
[Tesis Completa](#)

## Seguridad en el enrutamiento utilizando tecnología Blockchain

**Palabras clave:** Inter-domain routing security; BG; Blockchain

### Motivación

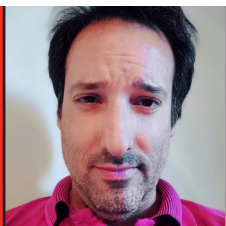
La tesis se centra en la seguridad del enrutamiento en Internet, específicamente en el protocolo BGP (Border Gateway Protocol), un problema persistente y crítico en el ámbito de las redes. Tradicionalmente, la seguridad de BGP ha sido un desafío debido a su dependencia de la confianza mutua entre sistemas autónomos (SAs), lo que la hace vulnerable a ataques como el Route Hijacking. La tesis busca abordar este problema desde una nueva perspectiva, utilizando la tecnología blockchain, que no requiere de la confianza mutua, ofreciendo así una oportunidad para mejorar significativamente la seguridad del enrutamiento de Internet.

### Aportes de la tesis

La investigación propone un modelo menos centralizado basado en blockchain para la asignación y delegación de recursos de Internet y la seguridad del ruteo externo BGP. Este modelo utiliza la red Ethereum Sepolia para almacenar información de ruteo, facilitando un sistema de enrutamiento seguro entre SAs. La tesis destaca la similitud entre los prefijos IP y las criptomonedas en términos de asignación y división, resaltando la aplicabilidad de blockchain en este contexto. Se implementa un prototipo funcional que incluye contratos inteligentes para la gestión de recursos de Internet, ofreciendo un modelo de confianza flexible, administración simplificada, alta auditabilidad y consenso distribuido sin certificados digitales ni control centralizado.

### Líneas de I/D futuras

La tesis abre varias líneas para investigaciones futuras en el campo de la seguridad del enrutamiento y la aplicación de la tecnología blockchain. Esto incluye la exploración de modelos alternativos y mejoras en el sistema propuesto para incrementar la eficiencia y efectividad en la prevención de ataques de Route Hijacking y similares. También sugiere el potencial de expandir este enfoque a otros aspectos de la seguridad en redes y la infraestructura de Internet, proponiendo un campo vasto para el desarrollo y la innovación tecnológica en sistemas descentralizados y seguros de gestión de recursos en Internet.



english

**Advisor**

Dra. Patricia Bazán

**Codirector**

Mg. Nicolás del Río

**Scientist Consultant**

Ing. Miguel Ángel Morandi

**Thesis approval date**

October 17, 2023

**SEDICI**

[Full Text](#)

## Routing security using Blockchain technology

**Keywords:** Inter-domain routing security; BGP; Blockchain

### Motivation

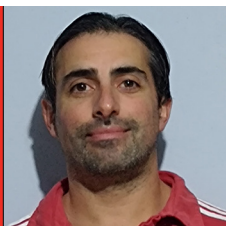
The thesis focuses on the security of Internet routing, specifically on the Border Gateway Protocol (BGP), a persistent and critical problem in the field of networks. Traditionally, the security of BGP has been a challenge due to its reliance on mutual trust between autonomous systems (ASs), making it vulnerable to attacks such as Route Hijacking. The thesis seeks to address this problem from a new perspective, using blockchain technology, which does not require mutual trust, thus offering an opportunity to significantly improve the security of Internet routing.

### Thesis contributions

The research proposes a less centralized blockchain-based model for the allocation and delegation of Internet resources and the security of external BGP routing. This model uses the Ethereum Sepolia network to store routing information, facilitating a secure routing system among Autonomous Systems (ASs). The thesis highlights the similarity between IP prefixes and cryptocurrencies in terms of allocation and division, emphasizing the applicability of blockchain in this context. A functional prototype is implemented, including smart contracts for the management of Internet resources. This offers a flexible trust model, simplified administration, high auditability, and distributed consensus without digital certificates or centralized control.

### Future Research Lines

The thesis opens several lines for future research in the field of routing security and the application of blockchain technology. This includes exploring alternative models and improvements to the proposed system to increase efficiency and effectiveness in preventing Route Hijacking attacks and similar threats. It also suggests the potential to expand this approach to other aspects of network security and Internet infrastructure, proposing a vast field for development and technological innovation in decentralized and secure systems for Internet resource management.



**Director**  
Mg. Alejandro Héctor González

**Codirector**  
Dra. María Alejandra Zangara

**Fecha de aprobación**  
16 de junio de 2023

**SEDICI**  
[Tesis Completa](#)

## Laboratorio virtual de electrónica básica para alumnos universitarios dentro de aula extendida

**Palabras clave:** Laboratorio; Virtual; Reutilizable; Competencias; Aula extendida; Simulación; Evaluación; Jurados

### Motivación

La formación práctica ocupa un lugar relevante dentro de las exigencias docentes, sobre todo con una importancia marcada en las carreras tecnológicas. La metodología habitual necesita la presencia física del alumno en el laboratorio para poder realizar las prácticas y el número de sesiones de prácticas libres está limitado por el tiempo de permanencia en el laboratorio. Debe sumarse a eso el inconveniente que presenta el ejercicio de la docencia en prácticas en laboratorios con equipamiento específico es que el trabajo previo y la preparación del ensayo que el alumno puede hacer fuera del laboratorio se ve limitado al número de horas que el laboratorio puede estar abierto.

La esencia de este proyecto es analizar los conocimientos que aportará el uso de un laboratorio virtual a la hora de realizar prácticas de estudios de circuitos eléctricos.

La práctica de laboratorio es una potente estrategia pedagógica para la construcción de competencias procedimentales y por este motivo es utilizada en una gran variedad de programas académicos, usualmente sincronizada con su asignatura teórica correspondiente. El trabajo de Infante Jiméñez aborda la importancia los laboratorios presenciales en las carreras de ciencias aplicadas por ejemplo Ingeniería, Fisicoquímica, Química, etc., sin embargo, a medida que los modelos educativos se han vuelto más flexibles y enfocados a competencias, la inclusión de las tic ha cambiado radicalmente el concepto de espacio físico; esto ha hecho patente una serie de limitaciones pues a pesar de la enorme importancia que éste tiene para el aprendizaje, no puede ofrecer la versatilidad idónea que se necesita en la actualidad. También es un hecho que la práctica de laboratorio presenta elevados tiempos de respuesta, los cuales tienden a bajar su productividad.

Se considera que el laboratorio remoto dentro del aula vir-

tual “ha sido un recurso importante para posibilitar la enseñanza técnica en la modalidad e-learning ya que ha posibilitado la realización de prácticas, a la vez que mantenía las características esenciales de la modalidad a distancia en cuanto a flexibilidad de horarios y espacio” del trabajo de (Albert Gómez et al) como conclusión del trabajo. Tal arribo de conclusión nuevamente afirma que los laboratorios remotos y virtuales son una excelente herramienta para la construcción del aprendizaje.

La utilización de herramientas de laboratorios virtuales se considera que generará una interacción alumno tutor a través de un entorno virtual formal de enseñanza. La tarea del tutor consistirá en orientar y reorientar los procesos de comprensión y de mediar pedagógicamente a través de las actividades auto estructuradas y progresivas del estudiante a distancia.

El objetivo general de este trabajo es evaluar la efectividad de realizar prácticas de electrónica básica con laboratorios virtuales dentro de un aula extendida.

Con el uso de metodologías de trabajo virtuales se puede lograr que un gran número de alumnos tenga acceso a los contenidos prácticos para poder familiarizarse con instrumentos y componentes de electrónica básica.

En el tema del trabajo se realizan prácticas de circuitos eléctricos de forma virtual, para posteriormente realizarlas dentro de un laboratorio físico donde se llevarán adelante mediciones con componentes eléctricos. Un aula extendida virtual puede aportar muchas ventajas en la docencia de asignaturas técnicas, pero también resulta de importancia evaluar el impacto que tiene sobre el alumnado.

Los objetivos particulares son:

- Ampliar el alcance de los trabajos prácticos propuestos en las clases de explicación práctica.

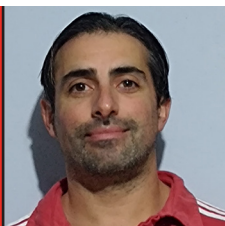
- Profundizar el análisis de los problemas planteados, llevando las experiencias a mediciones de situaciones reales.
- Analizar los indicadores de expectativas, estimulación, utilidad del laboratorio, uso del laboratorio remoto.
- Proponer orientaciones de uso del laboratorio virtual en el aula de electrónica.

### **Aportes de la tesis**

- Se realizó una serie de laboratorios virtuales en función a los trabajos prácticos de las asignaturas.
- Se promovió el uso de simuladores como herramienta de aprendizaje.
- Se estableció que las prácticas de laboratorio favorecen los procesos cognitivos.
- Se fomentó la comprobación de resultados al solucionar problemas utilizando simuladores.
- Se fomentó el uso de simuladores para la realización de trabajos en menor tiempo por parte de los alumnos.
- Se trabajó con herramientas de diseño multimedia, como ser diseño web, grabación de videos, edición de fotos.
- Se realizaron las experiencias de laboratorio verificando que los desarrollos sean acordes a los niveles de expertos necesarios.
- Se integraron los laboratorios virtuales en las distintas asignaturas que poseen contenidos de análisis de circuitos eléctricos.
- Se compartieron los contenidos desarrollados en cátedras Electrotecnia y Electrónica y Dispositivos Electrónicos, ambas dos correspondientes a la carrera Ingeniería en Computación (UNLP).
- Se modelaron situaciones reales, al hacer posible el logro de determinados objetivos educativos en cursos en los que se puedan aplicar.
- Se detectó la motivación por parte de los estudiantes ante este tipo de material.
- Se favoreció el proceso de evaluación, ya que los simuladores ofrecen una constante retroalimentación sobre las actuaciones de los alumnos, corrigiendo de forma inmediata los posibles errores.
- Se promovió la innovación con nuevas tecnologías.

### **Líneas de I/D futuras**

- Profundizar la línea investigativa en relación con las competencias tecnológicas a fin de adaptar el aprendizaje y la evaluación al nuevo paradigma de la formación orientada a las competencias.
- Articular entre los trabajos prácticos presentados en la asignatura con los contenidos de los laboratorios presentados donde se abordan situaciones reales.
- Mejorar la calidad gráfica de algunos diseños de los laboratorios



**Advisor**

Mg. Alejandro Héctor González

**Codirector**

Dra. María Alejandra Zangara

**Thesis approval date**

June 16, 2023

**SEDICI**

[Full Text](#)

## **Basic electronics virtual laboratory for university students in an extended classroom**

**Keywords:** Laboratory; Virtual; Reusable; Skills; Extended classroom; Simulation; Evaluation; Juries

### **Motivation**

Practical training occupies a relevant place within the teaching requirements, especially with a marked importance in technological careers. The usual methodology requires the physical presence of the student in the laboratory to be able to carry out the practices and the number of free practice sessions is limited by the time spent at the laboratory site. In addition to this, the inconvenience of teaching practices in laboratories with specific equipment is that the previous work and preparation of the essay that the student can do outside the laboratory is limited to the number of hours that the laboratory can be available.

The essence of this project is to analyze the knowledge that the use of a virtual laboratory will provide when carrying out practical studies of electrical circuits. The practical laboratory strategy is a powerful pedagogy for the construction of procedural competences and for this reason it is used in a great variety of academic programs, a norm synchronized with its corresponding theoretical subject. The work of Infante Jiménez shows the importance of face-to-face laboratories in applied science careers, such as Engineering, Physical Chemistry, Chemistry, etc., however, as educational models have become more flexible and focused on competencies, the inclusion IT&C have radically changed the concept of physical space; This has revealed a series of limitations because, despite its enormous importance for learning, it cannot offer the ideal versatility that is currently needed. It is also a fact that laboratory practice presents high response times, which tend to lower your productivity.

It is considered that the remote laboratory within the virtual classroom “has been an important resource to

enable technical teaching in the e-learning modality since it has made it possible to carry out practices, while having the essential characteristics of the distance modality in regarding flexibility of schedules and space” of the work of Albert Gómez et al as a conclusion of the work. Such a conclusion affirms once again that remote and virtual laboratories are an excellent tool for the construction of learning.

The use of virtual laboratory tools is considered to generate student-tutor interaction through a formal virtual teaching environment. The task of the tutor will consist of guiding and reorienting the comprehension processes and of pedagogically mediating through the self-structured and progressive activities of the distance student.

The general objective of this work is to evaluate the effectiveness of carrying out basic electronics practices with virtual laboratories within an extended classroom. With the use of virtual work methodologies, it is possible to achieve that many students have access to practical content to become familiar with basic electronic instruments and components.

In the subject of the work, electrical circuit practices are carried out virtually, to later be carried out in a physical laboratory where measurements with electrical components will be carried out. A virtual extended classroom can provide many advantages in the teaching of technical subjects, but it is also important to assess the impact it has on students.

The specific objectives are:

- Expand the scope of the practical work proposed in the practical explanation classes.



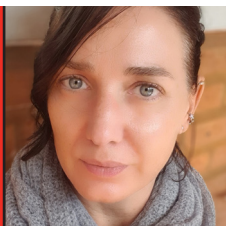
- Deepen the analysis of the problems raised, taking the experiences to measurements of real situations.
- Analyze the indicators of expectations, stimulation, utility of the laboratory, use of the remote laboratory.
- Propose guidelines for the use of the virtual laboratory in the electronics classroom.

### Thesis contributions

- A series of virtual laboratories were carried out based on the practical work of the subjects.
- The use of simulators as a learning tool was promoted.
- It was established that laboratory practices favor cognitive processes.
- The verification of results when solving problems using simulators was encouraged.
- The use of simulators was encouraged to carry out tasks in less time by the students.
- We worked with multimedia design tools, such as web design, video recording, photo editing.
- The laboratory experiences were carried out, verifying that the developments are in accordance with the necessary levels of experts.
- Virtual laboratories were integrated into the different subjects that have electrical circuit analysis content.
- The contents developed in Electrotechnics and Electronics and Electronic Devices chairs were shared, both corresponding to the Computer Engineering degree (UNLP).
- Real situations were modeled, making it possible to achieve certain educational objectives in courses in which they can be applied.
- Motivation on the part of the students was detected in the face of this type of material.
- The evaluation process was favored, since the simulators offer constant feedback on the actions of the students, immediately correcting possible errors.
- Innovation with new technologies was promoted.

### Future Research Lines

- Continue research in relation to technological competencies in order to adapt learning and assessment to the new paradigm of competency-oriented training.
- Make the link between the practical work presented in the subject with the contents of the laboratories presented where real situations are addressed.
- Improve the graphic quality of some laboratory designs.



**Directoras**  
Dra. Cecilia Verónica Sanz  
Dra. María Alejandra Zangara

**Fecha de aprobación**  
26 de septiembre de 2023

SEDICI  
[Tesis Completa](#)

## **Autorregulación del Aprendizaje, Estrategias de Participación y Rendimiento Académico. Un estudio de su relación en comunidades virtuales de indagación**

**Palabras clave:** autorregulación del aprendizaje; participación; rendimiento académico; foros de debate; pensamiento crítico; metodología de evaluación

### **Motivación**

En las últimas décadas, las instituciones educativas, en particular las de nivel superior, han sido testigos y partícipes de múltiples cambios impulsados por las tecnologías digitales. Quienes habitan estos espacios, con diferente grado de mediación tecnológica e hibridación espacio – temporal, debieron adoptar nuevas formas de enseñar, de aprender y de evaluar. En estos escenarios, de carácter flexible, accesible y colaborativo, con nuevas oportunidades de interacción, de acceso a recursos educativos y a la información en general, resulta vital la habilidad de los estudiantes para planificar, gestionar y controlar sus acciones de aprendizaje, con libertad de establecer sus propias metas y organizar sus tiempos. En otras palabras, la autonomía del estudiante permite hacer un mejor aprovechamiento de la estructura de la propuesta formativa y del diálogo que se genera entre los actores del hecho educativo. Y esto favorece el alcance de los logros académicos.

En esta tesis se abordan dos de estos tres elementos: la autonomía en términos de autorregulación del aprendizaje (ARA), y el diálogo entendido como una forma de interacción, relacional y discursiva, sociocultural y situada. Ambos constructos se estudian en vinculación con un tercero que preocupa en forma recurrente a las instituciones de educación superior y ha sido objeto de estudio permanente en la investigación educativa: el rendimiento académico. En particular, se intenta echar luz sobre estos tres elementos y las relaciones entre ellos en contextos formativos mediados tecnológicamente y entendidos como una comunidad virtual de indagación. El foro de debate es el espacio elegido para su abordaje. En particular, porque constituye un microcontexto tecnológico apropiado para la interacción y el desarrollo de tareas con fuerte demanda argumentativa orientadas al desarrollo del pensamiento crítico, la construcción de conocimiento y la resolución de problemas.

### **Objetivo general**

Identificar, analizar y evaluar relaciones significativas entre estrategias de ARA autoinformadas y manifiestas, características de participación y rendimiento académico en comunidades virtuales de indagación.

### **Objetivos específicos**

1. Investigar el constructo de ARA, los procesos involucrados, modelos y teorías existentes, e identificar su trascendencia en relación a la participación y el alcance de los logros académicos en contextos educativos mediados por tecnología digital.
2. Describir antecedentes de metodologías de evaluación de ARA en procesos formativos mediados por tecnología digital.
3. Investigar el modelo de comunidad de indagación, sus fundamentos y orientaciones, y posicionar a los foros virtuales como un microcontexto tecnológico adecuado para su estudio e indagación.
4. Describir antecedentes de metodologías e indicadores de seguimiento y evaluación de foros virtuales en los ámbitos educativo y tecnológico.
5. Definir una metodología para identificar, analizar y evaluar estrategias de ARA, características de participación, y rendimiento académico en una actividad de foro, a fin de definir los criterios e indicadores que permiten evaluar y valorar los procesos formativos que se llevan adelante en las comunidades virtuales de indagación.
6. Validar la metodología mediante un estudio de caso que permita analizar sus alcances y posibilidades e identificar eventuales mejoras a realizar.
7. Identificar, analizar y evaluar relaciones significativas entre estrategias de ARA, características de participación y rendimiento académico en una actividad de foro, a fin de definir las variables a considerar en el diseño, implementación y andamiaje de este tipo de tareas que favorezcan un buen desempeño por parte del estudiante.

## Aportes de la tesis

El desarrollo de esta tesis ha permitido generar un conjunto de aportes generales que se resumen a continuación:

- 1.** Una revisión teórica de los constructos de ARA, interacción y participación, y rendimiento académico con base en el análisis de literatura de relevancia científica y académica. De esta revisión se derivan definiciones y modelos explicativos de cada uno de los ejes temáticos mencionados.
- 2.** Una descripción del estado del arte, con base en una revisión sistemática de la literatura, de metodologías empleadas para evaluar ARA y rendimiento académico en procesos formativos mediados por tecnología digital, así como la participación en foros de debate.
- 3.** Una caracterización de los foros de debate académicos como un espacio virtual facilitador de la interacción, con posibilidades para promover el pensamiento crítico.
- 4.** El diseño, desarrollo y validación de una metodología denominada APReNDa, con base en el modelo de Diseño Centrado en la Evidencia (DCE), las revisiones y definiciones antes mencionadas y los estados del arte vinculados. APReNDa incluye múltiples instrumentos e indicadores destinados a identificar y evaluar el uso de estrategias de ARA, características de participación, y rendimiento académico en actividades destinadas a la construcción de conocimiento y el desarrollo del pensamiento crítico, implementadas en foros de debate en cursos mediados por tecnología digital. Además, sugiere un conjunto de técnicas analíticas orientadas a la identificación de relaciones entre los constructos mencionados. APReNDa puede emplearse, en forma total o parcial, por docentes e investigadores interesados en el seguimiento y evaluación de estos constructos o sus relaciones.
- 5.** La identificación de un conjunto de variables significativas a considerar en el diseño, implementación y andamiaje de tareas basadas en intercambios asincrónicos y escritos. Éstas derivan en un conjunto de recomendaciones orientadas a favorecer un buen desempeño, en términos de participación y rendimiento académico, por parte del estudiante.

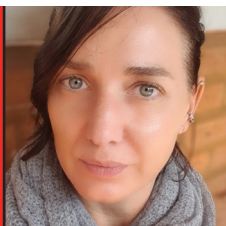
## Líneas de I/D futuras

Las conclusiones obtenidas en el proceso de investigación asociado a esta tesis derivan en una serie de trabajos futuros orientados a la mejora y ampliación de la metodología propuesta y a la validación externa de los resultados alcanzados. Asimismo, vislumbran algunas áreas de vacancia que podrían abordarse en investigaciones futuras. En particular:

- 1.** Expandir el estudio de campo en el marco de otros espacios formativos mediados que incorporen foros de debate académicos orientados a la construcción de conocimiento, el desarrollo de pensamiento crítico y la resolución de problemas, a fin de atender a la validez externa de los resultados alcanzados en la tesis.
- 2.** Someter los instrumentos diseñados ad hoc para APRe-

nda a pruebas alternativas que permitan reforzar la validez garantizada por el modelo DCE empleado en su diseño.

- 3.** Modificar o ampliar APReNDa para contemplar los aspectos identificados como puntos de mejora en el estudio de caso. Esto permitiría obtener condiciones necesarias para indagar con mayor fidelidad los constructos involucrados, ampliar el espectro de estrategias autorregulatorias y de participación indagadas, e identificar nuevas relaciones entre ellas y el rendimiento académico.
- 4.** Considerar procesos asociados con la correulación del aprendizaje y la regulación socialmente compartida para incorporar nuevas lentes con las que estudiar los procesos de intercambio y construcción grupal en las actividades de foro.
- 5.** Desarrollar un algoritmo automático que permita conformar equipos de trabajo atendiendo a criterios de grupalidad y que incorpore las variables identificadas en este trabajo como catalizadores del desempeño grupal.
- 6.** Avanzar hacia el desarrollo de protocolos automáticos orientados al análisis de contenido de los aportes en los foros de debate. Esto permitiría, además de acelerar y simplificar el procesamiento manual, demostrar las posibilidades del aprendizaje automático y del procesamiento de lenguaje natural para evaluar el aprendizaje y la colaboración en estos espacios de intercambio. Estos temas se pretenden abordar en una futura tesis doctoral.



**Advisors**

Dra. Cecilia Verónica Sanz  
Dra. María Alejandra Zangara

**Thesis approval date**

September 26, 2023

SEDICI

[Full Text](#)

**Learning self-regulation,  
participation strategies and  
academic performance.  
A study on how they relate  
in virtual communities of inquiry**

**Keywords:** self-regulation of learning; participation; academic performance; discussion forums; critical thinking; assessment methodology

**Motivation**

In recent decades, educational institutions, particularly higher education institutions, have witnessed and participated in multiple changes driven by digital technologies. Those who inhabit these spaces, with different degrees of technological mediation and spatio-temporal hybridisation, have had to adopt new ways of teaching, learning and assessing. In these flexible, accessible and collaborative scenarios, with new opportunities for interaction, access to educational resources and information in general, the ability of students to plan, manage and control their learning actions, with the freedom to set their own goals and organise their own time, is vital. In other words, the autonomy of students allows them to make better use of the structure of the educational proposal and of the dialogue that is generated between the actors involved in the educational process. And this favours the attainment of academic achievements.

This thesis addresses two of these three elements: autonomy in terms of self-regulation of learning (SRL), and dialogue understood as a form of interaction, relational and discursive, sociocultural and situated. Both constructs are studied in connection with a third one that is of recurrent concern to higher education institutions and has been a permanent object of study in educational research: academic performance. In particular, an attempt is made to shed light on these three elements and the relationships between them in technologically mediated learning contexts understood as a virtual community of inquiry. The discussion forum is the space chosen for its approach. In particular, because it constitutes an appropriate technological micro-context for interaction and the development of tasks with strong argumentative demands oriented towards the development of critical thinking, the construction of knowledge and problem solving.

**Main objective**

To identify, analyse and evaluate significant relationships between self-reported and overt SRL strategies, participation characteristics and academic performance in virtual communities of inquiry.

**Specific objectives**

1. To investigate the construct of SRL, the processes involved, existing models and theories, and to identify its significance in relation to participation and academic achievement in digitally mediated educational contexts.
2. To describe the background of methodologies for the assessment of SRL in educational processes mediated by digital technology.
3. To investigate the community of inquiry model, its foundations and orientations, and to position virtual forums as a suitable technological micro-context for its study and inquiry.
4. To describe background information on methodologies and indicators for monitoring and evaluating virtual forums in the educational and technological spheres.
5. To define a methodology to identify, analyse and evaluate SRL strategies, participation characteristics, and academic performance in a forum activity, in order to define the criteria and indicators that allow the evaluation and assessment of the educational processes that are carried out in virtual communities of inquiry.
6. To validate the methodology by means of a case study to analyse its scope and possibilities and identify possible improvements to be made.
7. To identify, analyse and evaluate significant relationships between SRL strategies, participation characteristics and academic performance in a forum activity, in order to define the variables to be considered in the design, implementation and scaffolding of this kind of tasks that favour good student performance.

## Thesis contributions

The development of this thesis has allowed us to generate a set of general contributions that are summarised below:

- 1.** A theoretical review of the constructs of SRL, interaction and participation, and academic performance based on the analysis of scientifically and academically relevant literature. From this review, definitions and explanatory models are derived for each of the aforementioned thematic axes.
- 2.** A description of the state of the art, based on a systematic review of the literature, of methodologies used to evaluate SRL and academic performance in educational processes mediated by digital technology, as well as participation in discussion forums.
- 3.** A characterisation of academic discussion forums as a virtual space that facilitates interaction, with possibilities for promoting critical thinking.
- 4.** The design, development and validation of a methodology called APReNDa, based on the Evidence-Centred Design (ECD) model, the aforementioned reviews and definitions and the related state of the art. APReNDa includes multiple instruments and indicators aimed at identifying and assessing the use of SRL strategies, participation characteristics, and academic performance in activities aimed at knowledge construction and the development of critical thinking, implemented in discussion forums in digitally mediated courses. In addition, it suggests a set of analytical techniques aimed at identifying relationships between the aforementioned constructs. APReNDa can be used, in whole or in part, by teachers and researchers interested in monitoring and evaluating these constructs or their relationships.
- 5.** The identification of a set of significant variables to be considered in the design, implementation and scaffolding of tasks based on asynchronous and written exchanges. These derive in a set of recommendations oriented to favour a good performance, in terms of participation and academic performance, on the part of the student.

## Future Research Lines

The conclusions obtained in the research process associated with this thesis lead to a series of future works aimed at the improvement and extension of the proposed methodology and the external validation of the results achieved. They also provide a glimpse of some areas of vacancy that could be addressed in future research. In particular:

- 1.** To expand the field study in the framework of other mediated formative spaces that incorporate forums for academic debate oriented towards knowledge construction, the development of critical thinking and problem solving,

in order to address the external validity of the results achieved in the thesis.

- 2.** To subject the instruments designed ad hoc for APReNDa to alternative tests to reinforce the validity guaranteed by the ECD model used in their design.
- 3.** To modify or extend APReNDa to contemplate the aspects identified as points for improvement in the case study. This would make it possible to obtain the necessary conditions to investigate the constructs involved with greater accuracy, to broaden the spectrum of self-regulatory and participation strategies investigated, and to identify new relationships between them and academic performance.
- 4.** To consider processes associated with co-regulation of learning and socially shared regulation in order to incorporate new lenses with which to study the processes of exchange and group construction in forum activities.
- 5.** To develop an automatic algorithm that allows us to form work teams according to group criteria and that incorporates the variables identified in this work as catalysts of group performance.
- 6.** To move towards the development of automatic protocols oriented towards the content analysis of the contributions in the discussion forums. In addition to speeding up and simplifying manual processing, this would demonstrate the possibilities of machine learning and natural language processing to evaluate learning and collaboration in these exchange spaces. These issues are intended to be addressed in a future doctoral thesis.



**03**

ESPECIALIZACIONES

**INTELIGENCIA DE DATOS  
ORIENTADA A BIG DATA**

**INGENIERÍA DE SOFTWARE**

**TECNOLOGÍA INFORMÁTICA  
APLICADA EN EDUCACIÓN**

## ESPECIALIZACIÓN INTELIGENCIA DE DATOS ORIENTADA A BIG DATA



**Esp. Alejo Hernández**

e-mail

homocadratus@gmail.com

**Director**  
Dr. Aurelio Fernandez Bariviera

**Codirector**  
Dr. Ricardo Di Pasquale

**Fecha de aprobación**  
16 de febrero de 2023

**SEDICI**  
[Tesis Completa](#)

## Bases de datos de series temporales y métodos estadísticos para la predicción de rendimientos de ETH

**Palabras clave:** Base de datos; series temporales; ETH

### Motivación

Las bases de datos forman parte de nuestra vida diaria, aunque muchos de nosotros probablemente no seamos conscientes de este hecho. Esto es fácil de entender si tenemos en cuenta que la mayoría de las transacciones electrónicas nos ponen en contacto con una de ellas y que este tipo de transacciones se han vuelto ubicuas, dado el nivel de penetración que poseen los teléfonos celulares y el crecimiento exponencial que han experimentado los dispositivos conectados a internet (IoT, IoE).

Este uso intensivo de las bases de datos, combinado con la vastedad de escenarios en los cuales se ven involucradas, ha propiciado un proceso evolutivo que comenzó en la década del '60 y continúa hasta nuestros días. En este camino, varios tipos de bases de datos fueron creadas: relacionales, no-SQL, de almacenamiento en memoria, de grafos, geoespaciales y por último, las específicas para series temporales (BDSTs).

Por otra parte, durante los últimos diez años la adopción de las criptomonedas como instrumento de inversión y/o de especulación fomentó el desarrollo de un mercado de negociación deslocalizado, de muy alta frecuencia y sin interrupciones significativas que, al día de hoy, se presenta como una alternativa viable para el manejo de divisas y el pago de bienes y servicios. Si bien en términos académicos, la mayoría de los trabajos publicados utilizan datos con frecuencias diarias, esta tendencia está cambiando, puesto que los investigadores han visto la importancia de extraer información de datos en más alta frecuencia para capturar el comportamiento de este mercado altamente dinámico.

Estas características particulares que presenta el mercado de criptomonedas se traducen en un flujo de datos continuo y de alta frecuencia, modulado mayoritariamente por operaciones de trading automatizadas. Esto evidencia la ne-

cesidad de contar, por un lado, con un sistema de almacenamiento y procesamiento de datos acorde con el volumen y velocidad de generación involucrados y por otro, con una serie de algoritmos robustos que permitan predecir los retornos de los activos de forma precisa y confiable. A fin de explorar más en profundidad estos aspectos, este trabajo se propone implementar un pipeline de datos soportado por una BDST a fin de realizar predicciones de rendimiento de activos usando métodos estadísticos. Más concretamente, se implementa la BDST InfluxDB para predecir los retornos de Ethereum (ETH), la segunda criptomoneda en términos de capitalización de mercado. Desde el punto de vista del manejo de datos, se decidió trabajar con la BDST InfluxDB puesto que la misma cuenta con una suite de herramientas muy completa para gestionar los procesos de captura, almacenamiento, consulta y visualización de la información. Por otra parte, en lo relativo al modelado estadístico de los datos, se utilizó el paquete Statsmodels de Python que permite realizar la predicción de los retornos aplicando un modelo ARMA con intervalos de confianza corregidos mediante un modelo GARCH.

### Aportes del Trabajo Final Integrador

En este trabajo se implementó un pipeline de datos soportado por una base de datos de series temporales (BDST) que permite predecir el rendimiento de criptoactivos usando métodos estadísticos. Puntualmente, se descargaron manualmente datos de trading diarios relativos a la criptomoneda Ethereum (ETH) desde la página CoinmarketCap, se pre-procesaron e insertaron en la BDST InfluxDB y se hicieron consultas sobre la misma para entrenar un modelo estadístico de tipo ARMA con intervalos de confianza corregidos



empleando GARCH, utilizado para predecir los rendimientos diarios de ETH.

Durante este proceso, se exploró el impacto actual que tienen las BDST desde un punto de vista conceptual y se evaluaron las capacidades de las mismas al instalar y utilizar la suite de InfluxDB. Este último proceso demandó para su éxito el desarrollo de varias competencias, a saber: el manejo de configuraciones internas del sistema operativo utilizado (host), así como de la propia InfluxDB; el estudio del modelo de datos empleado por la BDST; el aprendizaje del protocolo de línea utilizado por la BDST para escribir los datos; el aprendizaje del lenguaje de consulta Flux para interactuar con InfluxDB.

Por otra parte, tanto el repaso de conceptos estadísticos así como también la formalización de los modelos fueron determinantes para la correcta ejecución de las acciones de modelado y para la interpretación precisa de los pronósticos obtenidos. En relación con el proceso de modelado de los datos de trading diarios de ETH y la posterior predicción de los rendimientos, cabe destacar que los procedimientos empleados son acordes a las mejores prácticas de la disciplina. Por otra parte, si bien el desarrollo de este trabajo involucró un enfoque estadístico para realizar las estimaciones de los rendimientos, es dado mencionar que existen varias técnicas de Machine Learning que podrían haberse empleado para obtener las mismas. Considerando que el trabajo en finanzas cuantitativas se basa en la predicción de series temporales, que las mismas son muchas y de las más variadas características y que los métodos disponibles para realizar los pronósticos son también múltiples, se comprende la importancia que tiene contar con un pipeline lo suficientemente versátil como para simplificar el proceso de gestión de las series temporales, la metadata asociada y el modelado por distintas vías.

### Líneas de I/D futuras

Nos proponemos explorar a futuro mejoras en el pipeline, como por ejemplo la automatización de la captura de los datos financieros vía consulta directa a APIs de sitios de intercambio (Binance, CoinmarketCap, etc.) así como también la implementación de algoritmos de Machine Learning más complejos que los modelos estadísticos básicos aplicados en el presente trabajo.

ESPECIALIZACIÓN  
INTELIGENCIA DE DATOS  
ORIENTADA A BIG DATA



**Esp. Alejo Hernández**

**e-mail**

homocuatratus@gmail.com

**Advisor**  
Dr. Aurelio Fernandez Bariviera

**Codirector**  
Dr. Ricardo Di Pasquale

**Thesis approval date**  
February 16, 2023

**SEDICI**  
[Full Text](#)

## **Database of time-series and statistical methods to predict ETH performance**

**Keywords:** Databases; time-series; ETH

### **Motivation**

Databases are part of our daily life; however, we are not all probably aware of their existence. This is easy to understand if we consider that most of our electronic exchanges contact us with them. These transactions are omnipresent, for example, in our mobile phones and in the exponential growth of all the devices that connect to Internet (IoT, IoE).

On the one hand, this intensive use of databases, in combination with the vast of sceneries where they are involved, lead to an evolving process that has begun in the 60s and it is still present. Thus, several types of databases, such as relational, NoSQL, in-memory, graph, geospatial and time series (TSBDs) have been created.

On the other hand, during the last ten years, the incorporation of cryptocurrency as a means of inversion and/or speculation encouraged a delocalized negotiation market, highly frequent and with almost no interruptions which is a viable option today to invest money and to pay goods or services. Academically speaking, most of the publications use daily data, but this trend is changing because researchers consider important to extract information more frequently to capture the behaviour of this highly dynamic market.

The particular features of the cryptocurrency market are translated in a continuous and highly frequent data flow, in special with automated trading operations. So, it is vital to have, on the one hand, a storage and data processing system that fit the volume and speed of generation involved and, on the other hand, a series of robust algorithms that enable predicting returns of actives in a precise and faithful manner. In order to deep into all these

aspects, this work aims at implementing a data pipeline supported by a TSBD to predict active performance by means of statistical methods. In concrete, an InfluxDB TSBD is implemented to predict Ethereum (ETH) returns, the second cryptocurrency in market capitalization. From the point of view of data management, we decided to work with TSBD InfluxDB because it has a very complete suite of tools to manage the processes of capture, storage, consulting and data visualization. On the other hand, in terms of data statistical modelling, we used the pack of Python Statsmodels which enables prediction of returns applying an ARMA model with confidence intervals checked by means of a GARCH model.

### **Thesis Final Work contributions**

A data pipeline supported by a time series database (TSDB) was here implemented to predict the cryptocurrency performance by means of statistical methods. In particular, daily trading data were manually downloaded as regards Ethereum (ETH) from the website of CoinmarketCap; InfluxDB was processed and inserted in the TSDB, and some consulting about that cryptocurrency was done to train an ARMA statistic model with modified confidence intervals using GARCH to predict the ETH daily performance.

The current impact of TSDB from a conceptual perspective as well as the skills they have to install and use InfluxDB suite were evaluated during this process. But it was necessary to develop several skills to succeed: internal configuration management of the operating system employed (host) and the InfluxDB; the study of the data model used by the TSDB; learning the protocol used by

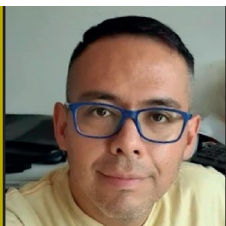
the TSDB to write data, and learning the Flux consulting language to interact with InfluxDB.

On the other hand, the review of statistical concepts and the formalization of models were decisive to correctly execute modelling actions and precisely interpret the obtained prediction. As regards data modelling process of ETH daily trading and the later prediction of its performance, it is worth highlighting that the employed procedures are consistent with the best practices of the discipline. Despite the fact that this work development includes a statistical approach to estimate performance, it is worth mentioning that there exist other Machine Learning techniques that could have been used to get to the same point. As working on quantitative finance means predicting time series, and that there are several and with multiple features, and that there exist multiple available methods for prediction, it is really important to work with a versatile pipeline so as to simplify the process of time series management, the associated metadata and the modelling processes through different ways.

### **Future Research Lines**

In a near future, our aim is to improve the pipeline, i.e. to automate financial data capture by means of a direct consulting in APIs of exchange (Binance, CoinmarketCap, etc.) and also to implement some Machine Learning algorithms which are more complex than the basic statistical models used in this work.

## ESPECIALIZACIÓN INTELIGENCIA DE DATOS ORIENTADA A BIG DATA



**Esp. Hugo Manuel Fajardo**

e-mail

hugofajardo1@gmail.com

español

**Director**  
Dr. Waldo Hasperué

**Fecha de aprobación**  
9 de mayo de 2023

**SEDICI**  
[Tesis Completa](#)

## **Estudio Comparativo entre Apache Spark y Apache Flink en el procesamiento de streaming en entornos Big Data**

**Palabras Clave:** Streaming de Datos; Procesamiento de Flujos; Procesamiento Distribuido de Flujos de Datos; Apache Spark; Apache Flink; Apache Kafka

### **Motivación**

El procesamiento de streaming resulta beneficioso en la mayoría de las situaciones en las que se generan datos nuevos y dinámicos de forma continua. Es apto para la mayoría de los sectores y casos de uso de Big Data. Por lo general, las empresas comienzan con aplicaciones sencillas, que realizan recopilación de registros, e implementan cálculos de valores medios, mínimos y máximos. Más adelante, estas aplicaciones evolucionan y se pasa a un procesamiento más sofisticado, casi en tiempo real. En un principio, las aplicaciones pueden procesar transmisiones de datos para producir informes básicos y realizar acciones sencillas como respuesta, por ejemplo, emitir alertas cuando las medidas clave superan ciertos umbrales. Con el tiempo, dichas aplicaciones realizan un análisis de datos más sofisticado, como la aplicación de algoritmos de aprendizaje automático y la extracción de información más exhaustiva a partir de los datos. Posteriormente, se aplican complejos algoritmos de procesamiento de transmisiones y eventos, como intervalos de tiempo decrecientes para encontrar, por ejemplo, las películas populares más recientes, lo que enriquece aún más la información. Apache Spark es un motor de análisis para el procesamiento de datos a gran escala, desarrollado específicamente para el procesamiento distribuido por lotes, y con el tiempo se adaptó y optimizó para el procesamiento de streaming. Flink es un motor de procesamiento distribuido diseñado y construido específicamente para el trabajo con flujos de datos ilimitados y acotados.

El objetivo del presente trabajo es realizar una comparación entre los dos frameworks previamente mencionados. El estudio comparativo se realizará desde varios puntos de vista, entre ellos:

- Facilidad de instalación y despliegue.
- Fuentes admitidas para el intercambio de datos.
- Lenguajes de programación soportados.
- Documentación disponible.
- Evaluación de rendimiento.

Para realizar la comparación entre estas herramientas se utilizará el lenguaje de desarrollo Python, ya que el mismo soporta

el trabajo tanto en Spark como en Flink, y a su vez es uno de los lenguajes de programación más utilizados en la actualidad. La comparación entre los frameworks requiere el desarrollo de dos aplicaciones para el tratamiento del flujo de datos, ambas resolviendo el mismo problema. Una aplicación realizará el procesamiento del flujo de datos en Apache Spark, mientras que la otra realizará la misma tarea en Apache Flink. El problema puntual será la implementación de una técnica de minería de datos, mientras que el streaming será construido de manera sintética simulando un flujo real.

### **Aportes del Trabajo Final Integrador**

Con frecuencia los trabajos de comparación entre herramientas tecnológicas se centran en el comparar el aspecto cuantitativo, generalmente vinculado a evaluación de rendimiento. En el presente trabajo se buscó, además, incluir en la comparación aspectos cualitativos entre ellos:

- Facilidad de instalación y despliegue.
- Fuentes admitidas para el intercambio de datos.
- Lenguajes de programación soportados.
- Documentación disponible.

De esta forma los resultados de rendimiento obtenidos pueden complementarse con valiosa información que resulta importante cuando se necesita elegir una herramienta concreta de procesamiento de streaming.

### **Líneas de I/D futuras**

Una característica que Spark y Flink anuncian es la posibilidad de disponer de algoritmos avanzados de Machine Learning. Esto abre las puertas a trabajos futuros, trabajos que complementen el presente estudio comparativo con análisis de algoritmos de Machine Learning.

También puede ser un trabajo futuro posible, integrar pipelines de procesamiento de flujos y pipelines de procesamiento por lotes, ya que en muchos casos las necesidades de procesamiento de streaming deben convivir con casos de uso de procesamiento por lotes.



**Advisor**  
Dr. Hasperué Waldo  
**Thesis approval date**  
May 9, 2023  
**SEDICI**  
[Full Text](#)

## **A comparison between Apache Spark and Apache Flink and how they process streaming in Big Data environments**

**Keywords:** Data Streaming; Flux Processing; Data Flux Distributed Processing; Apache Spark; Apache Flink; Apache Kafka

### **Motivation**

The stream processing is mostly efficient in those situations where new and dynamic data are created continuously. It is useful for almost every sector and when Big Data is used. In general, companies start using simple applications which gather records and implement operations of medium, minimum and maximum values. Later, these applications evolve and a more sophisticated processing is reached in almost real time. Firstly, the applications can process data transmission to produce basic records and carry out simple actions, such as an alert if the key measures are higher. Then, those applications carry out a more sophisticated data analysis, for example, applying automatic learning algorithms and extracting more exhaustive information from them. Finally, some complex algorithms of transmission and event processing are applied, such as decreasing time intervals to find, for example, the latest most popular movies, which makes information even richer.

Apache Spark is an analytic engine for large-scale data processing; it was developed specifically for the distributed processing of clusters; and later, it was adapted and optimized for stream processing. Flink is a distributed processing engine designed and built specifically to work with unlimited and bounded data streams.

The aim of this present work is to make a comparison between these two frameworks mentioned herein above. This comparison is made from several points of view:

- Easy setup and display.
- Accepted sources for data exchange.
- Programming languages accepted.
- Available information.
- Performance assessment.

In order to make the comparison, we are going to use the Python development language because it supports this work in Spark as well as in Flink and, at the same time, it is one of the mostly used programming languages nowadays.

The comparison between frameworks requires the development of two applications to work with data stream, and both should solve the same problem. One application processes the data flux in Apache Spark, meanwhile the other does the same using Apache Flink. Our point is implementing a data mining technique while the streaming will be briefly built pretending a real flux.

### **Thesis Final Work contributions**

Frequently, comparisons between technological tools focus on the quantity as regards performance assessment. However, this work tries to include a qualitative comparison between them:

- Easy setup and display.
- Accepted sources for data exchange.
- Programming languages accepted.
- Available information.

So, the performance results can be added to some other valuable information at the time of choosing a particular stream processing tool.

### **Future Research Lines**

One feature announced by Spark and Flink is the possibility to use advanced Machine Learning algorithms. This opens to future works that may complement the present comparison using Machine Learning algorithm analysis.

Another future work may be the integration of flux processing pipelines and cluster processing pipelines because in many cases the stream processing needs should cohabit with cluster processing case uses.

## ESPECIALIZACIÓN INTELIGENCIA DE DATOS ORIENTADA A BIG DATA



**Esp. Mercedes Barrionuevo**

e-mail

mercedes.barrionuevo@gmail.com

español

**Directora**  
Dra. María Luján Ganuza

**Fecha de aprobación**  
26 de junio de 2023

**SEDICI**  
[Tesis Completa](#)

## **Análisis Visual de Datos Temporales Asociados a Criptomonedas**

**Palabras clave:** Visualización; Datos Temporales; Criptomonedas; Análisis de Datos Financieros

### **Motivación**

El objetivo general del trabajo integrador es contribuir al desarrollo de soluciones en torno al análisis visual de indicadores relevantes para el mercado de las criptomonedas. Se plantea el diseño de una herramienta de visualización para estos datos de características temporales y de las interacciones asociadas.

Con este objetivo, se establecen como objetivos específicos:

- El diseño de una herramienta de análisis visual con vistas coordinadas para datos temporales, específicamente datos financieros de criptomonedas.
- El diseño de las interacciones necesarias para el análisis de características como: volatilidad de las monedas, detección de patrones de comportamiento de interés, comparación entre distintas monedas, entre otras.
- El diseño de las interacciones específicas inteligentes para la detección y representación visual de repetición de patrones tanto para una moneda, como para patrones combinados que involucren más de una moneda.
- Implementación de un prototipo utilizando alguna de las herramientas configurables disponibles (Tableau, Power BI, entre otras).
- Validar los resultados obtenidos con un grupo de pares en congresos o revistas.

### **Aportes del Trabajo Final Integrador**

El aporte del trabajo es contribuir al desarrollo de soluciones en torno al análisis visual de indicadores relevantes para el mercado de las criptomonedas, como así también, ayudar a la comparación del comportamiento de diversas criptomonedas planteadas en un mismo dashboard.

### **Líneas de I/D futuras**

Como trabajo a futuro se propone realizar un estudio más exhaustivo de las distintas características de los gráficos de visualización financieros. Esto involucra la integración al equipo de trabajo de especialistas del dominio del problema como economistas, expertos en finanzas y/o criptomonedas.

También, se pretende realizar una implementación del prototipo en un lenguaje de más bajo nivel (Python, R, JavaScript) que permita, a través de librerías específicas de visualización de datos, incorporar todo el diseño original.

Por otra parte, es de interés realizar la automatización de la detección de patrones para implementar las interacciones inteligentes planteadas, que debido a las limitaciones propias de la herramienta configurable no pudieron implementarse en el presente trabajo.

Finalmente, se desea publicar en un futuro cercano, los resultados obtenidos en distintos congresos, workshops o eventos nacionales.



english

**Advisor**  
Dra. María Luján Ganuza

**Thesis approval date**  
June 16, 2023

**SEDICI**  
[Full Text](#)

## **Visual Analysis of Temporary Data Associated with Cryptocurrencies**

**Keywords:** Visualization; Temporary Data; Cryptocurrencies; Financial Data Analysis

### **Motivation**

The general objective of the integrating work is to contribute to the development of solutions around the visual analysis of relevant indicators for the cryptocurrency market. The design of a visualization tool for these data of temporal characteristics and associated interactions is proposed.

With this objective, the following are established as specific objectives:

- The design of a visual analysis tool with coordinated views for temporal data, specifically cryptocurrency financial data.
- The design of the necessary interactions for the analysis of characteristics such as: volatility of currencies, detection of patterns of behavior of interest, comparison between different currencies, among others.
- The design of specific intelligent interactions for the detection and visual representation of repeating patterns for both one coin and combined patterns involving more than one coin.
- Implementation of a prototype using one of the configurable tools available (Tableau, Power BI, among others).
- Validate the results obtained with a group of peers in congresses or journals.

### **Thesis contributions**

The contribution of the work is to contribute to the development of solutions around the visual analysis of relevant indicators for the cryptocurrency market, as well as to help compare the behavior of various cryptocurrencies raised on the same dashboard.

### **Future Research Lines**

As future work, it is proposed to carry out a more exhaustive study of the different characteristics of financial visualization graphs. This involves the integration of specialists in the problem domain such as economists, experts in finance and/or cryptocurrencies into the work team.

## ESPECIALIZACIÓN INGENIERÍA DE SOFTWARE



**Esp. Luciano Marrero**

**e-mail**

marreroluciano@gmail.com

español

**Director**  
Mg. Pablo Thomas  
**Fecha de aprobación**  
29 de mayo de 2023  
**SEDICI**  
[Tesis Completa](#)

## Un estudio comparativo de Bases de Datos NoSQL

**Palabras clave:** Base de Datos NoSQL; Bases de Datos Relacionales; Almacenamiento no Estructurado de Datos; Conceptos generales de Bases de Datos.

### Motivación

Las bases de datos relacionales han demostrado cubrir las necesidades de almacenamiento de información en una gran variedad de soluciones informáticas. No obstante, cuando estas bases de datos comenzaron a ser utilizadas, los esquemas y el crecimiento del volumen de datos era muy previsible en el tiempo. Actualmente, la realidad es muy distinta, existen aplicaciones que generan enormes volúmenes de información y de forma no estructurada. Ante esta realidad, se utilizan nuevas formas de almacenamiento denominadas bases de datos no relacionales o NoSQL. El objetivo de este trabajo final consiste en elaborar un análisis comparativo entre las diferentes categorías de almacenamiento no relacional de datos.

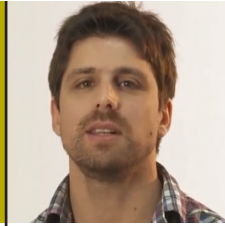
### Aportes del Trabajo Final Integrador

El aporte principal de este trabajo es brindar un conjunto de experimentos que permitan evaluar comparativamente los motores de Bases de Datos MySQL, MongoDB, Apache Cassandra, Neo4j, Redis y dos motores de Bases de Datos como servicio en la nube, Cloud Firestore y MongoDB Atlas. Exceptuando a MySQL, motor de Base de Datos relacional, los restantes son todos motores de Bases de Datos NoSQL. Además, el trabajo pretende ser una guía teórica inicial sobre base de datos en general.

### Líneas de I/D futuras

Ante la variedad de resultados obtenidos en el trabajo y la falta de prevalencia de un motor de Bases de Datos con el mejor rendimiento, como trabajo futuro se prevé plantear escenarios para realizar un escalamiento horizontal de los esquemas planteados y realizar una nueva evaluación en ese nuevo contexto. Además, se buscará incorporar nuevos experimentos y reforzar las existentes con alternativas de motores de Bases de Datos relacionales y NoSQL en diversos ambientes.





english

**Advisor**  
Mg. Pablo Thomas  
**Thesis approval date**  
May 29, 2029  
**SEDICI**  
[Full Text](#)

## **BD, NO and SQL: a comparison**

**Keywords:** NoSQL Database; Relational Databases; Unstructured Data Storage; General Database Concepts.

### **Motivation**

Relational databases have proven to meet information storage needs in a wide variety of computing solutions. However, when these databases began to be used, the schemas and the growth in the volume of data was predictable over time. Today, the reality is very different; there are applications that generate huge volumes of information in an unstructured way. Faced with this reality, new forms of storage are being used, called non-relational or NoSQL databases. The objective of this final work is to elaborate a comparative analysis between the different categories of non-relational data storage.

### **Thesis Final Work contributions**

The main contribution of this work is to provide a set of experiments that allow a comparative evaluation of database engines MySQL, MongoDB, Apache Cassandra, Neo4j, Redis and two Database-as-a-Service engines in the cloud, Cloud Firestore and MongoDB Atlas. With the exception of MySQL, a relational database engine, the others are all NoSQL database engines. In addition, the paper is intended as an initial theoretical guide to databases in general.

### **Future Research Lines**

Given the variety of results obtained in the work and the lack of prevalence of a database engine with the best performance, as future work we plan to propose scenarios to perform a horizontal scaling of the proposed schemes and perform a new evaluation in this new context. In addition, we will seek to incorporate new experiments and reinforce the existing ones with alternatives of relational and NoSQL database engines in different environments.

## ESPECIALIZACIÓN TECNOLOGÍA INFORMÁTICA APLICADA EN EDUCACIÓN



**Esp. Leandro Javier Castro**

e-mail  
leajcastro@gmail.com

español

**Directora**  
Dra. Verónica Artola

**Codirector**  
Mg. Gustavo Astudillo

**Fecha de aprobación**  
9 de Junio de 2023

**SEDICI**  
[Tesis Completa](#)

## **Programación tangible en el ámbito educativo. Análisis de experiencias**

**Palabras clave:** educación; programación tangible; interfaces de usuario tangible, estado del arte

### **Motivación**

Describir un estado del arte de la utilización de la programación tangible en el contexto educativo. Para ello se definieron los siguientes objetivos de investigación:

- Identificar las bases teóricas y/o enfoques que sustentan la programación tangible.
- Identificar distintas experiencias que utilizan programación tangible en contextos educativos en los últimos diez años.
- Definir un conjunto de criterios que permitan el análisis de las experiencias seleccionadas.
- Reconocer ventajas y desventajas de la utilización de la programación tangible en contextos educativos.
- Analizar las experiencias educativas que hacen uso de programación tangible.

### **Aportes del Trabajo Final Integrador**

Describir un estado del arte de la utilización de la programación tangible en el contexto educativo entre los años 2011 y 2021.

### **Líneas de I/D futuras**

En base a la investigación realizada se espera poder validar los criterios de evaluación definidos a través de la validación por juicio de expertos y su aplicación en nuevas experiencias. Esto permitirá estandarizar el uso de los criterios definidos. Asimismo, se pretende diseñar y desarrollar una TUI que se presente como una alternativa para incorporar las nociones básicas sobre programación en alumnas y alumnos que no tienen conocimientos previos sobre esta temática. Se considera de interés tener presente que la herramienta sea accesible y usable por personas con problemas motrices o

visuales. La implementación y prueba de la TUI permitirá, además, sumar resultados que den sustento a un enfoque de aprendizaje de la programación más intuitivo y usable a través del uso de programación tangible. Al mismo tiempo se pretende recabar información que sustente las ventajas y desventajas en el uso de una TUI en un primer acercamiento hacia la programación.

ESPECIALIZACIÓN  
TECNOLOGÍA INFORMÁTICA  
APLICADA EN EDUCACIÓN



**Esp. Leandro Javier Castro**

e-mail  
leajcastro@gmail.com

english

**Advisor**  
Dra. Verónica Artola

**Codirector**  
Mg. Gustavo Astudillo

**Thesis approval date**  
June 9, 2023

**SEDICI**  
[Full Text](#)

**Tangible programming  
in education.  
Analyzing experiences**

**Keywords:** education; tangible programming; tangible user interfaces, state Of art.

## Motivation

Describe a state-of-the-art of the use of tangible programming in the educational context. For this purpose, the following research objectives were defined:

- Identify the theoretical foundations and/or approaches that support tangible programming.
- Identify different experiences that utilize tangible programming in educational contexts over the past ten years.
- Define a set of criteria that allow the analysis of the selected experiences.
- Recognize the advantages and disadvantages of using tangible programming in educational contexts.
- Analyze educational experiences that make use of tangible programming.

## Thesis Final Work contributions

Describe a state-of-the-art of the use of tangible programming in the educational context between the years 2011 and 2021.

## Future Research Lines

Based on the research carried out, it is expected to be able to validate the evaluation criteria defined through validation by expert judgment and its application in new experiences. This will allow standardizing the use of the defined criteria. Likewise, it is intended to design and develop a TUI that is presented as an alternative to incorporate basic notions about programming in students who have no prior knowledge on this subject. It is considered of interest to keep in mind that the tool is accessible and usable by people with motor

or visual problems. The implementation and testing of the TUI will also allow adding results that support a more intuitive and usable programming learning approach through the use of tangible programming. At the same time, it is intended to collect information that supports the advantages and disadvantages of using a TUI in a first approach to programming.



# JURADOS DESIGNADOS 2023

Dr. Acosta Nelson (UNICEN - Argentina)  
Dr. Antonelli Leandro (UNLP - Argentina)  
Dr. Arias Figueroa Daniel (UNSA - Argentina)  
Mg. Astudillo Gustavo (UNLPam - Argentina)  
Mg. Banchoff Claudia (UNLP - Argentina)  
Mg. Bertone Rodolfo (UNLP - Argentina)  
Dr. Cuevas Alfredo Simón (CUJAE - Cuba)  
Lic. Dapoto Sebastián (UNLP - Argentina)  
Dr. Falappa Marcelo (UNS - Argentina)  
Dr. Frati Emmanuel (UNdeC-Argentina)  
Dr. Guerrero Roberto (UNSL - Argentina)  
Dr. Hasperué Waldo (UNLP - Argentina)  
Dr. Kuna Horacio (UNAM - Argentina)  
Dr. Lagares Antonio (UNLP - Argentina)  
Dr. Leguizamón Guillermo (UNSL - Argentina)  
Mg. Lliteras Alejandra (UNLP - Argentina)  
Dr. Micolini Orlando (UNC - Argentina)  
Dr. Mitaritonna Alejandro (UNDEF - Argentina)  
Dr. Moré Jerónimo (UNLP - Argentina)  
Dr. Olivas Varela José Ángel (UCLM - España)  
Dr. Piattini Mario (UCLM - España)  
Dr. Piccirilli Dario (UNLP - Argentina)  
Dr. Pousa Adrián (UNLP - Argentina)  
Dr. Quiroga Facundo Manuel (UNLP - Argentina)  
Dr. Riesco Daniel (UNSL - Argentina)

Dr. Rossi Gustavo (UNLP - Argentina)  
Dr. Rucci Enzo (UNLP - Argentina)  
Dr. Sosa Eduardo Omar (UNAM - Argentina)  
Dr. Suppi Boldrito Remo (UAB - España)  
Dra. Baldassarri Sandra (UNIZAR - España)  
Dra. Bazán Patricia (UNLP - Argentina)  
Dra. Britos Paola (UNRN - Argentina)  
Dra. Castro Silvia (UNSA - Argentina)  
Dra. De Giusti Laura (UNLP - Argentina)  
Dra. Genero Bocco Marcela (UCLM - España)  
Dra. Godoy Daniela (UNCPBA - Argentina)  
Dra. Herrera Susana (UNSE - Argentina)  
Dra. Lanzarini Laura (UNLP - Argentina)  
Dra. Luque Mónica (UCC - Argentina)  
Dra. Marín Beatriz (UPV - España)  
Dra. Mon Alicia (ITBA - Argentina)  
Dra. Olcoz Katzalin (UCM - España)  
Dra. Otero Rita (UNICEN - Argentina)  
Mg. Ramón Hugo (UNNOBA - Argentina)  
Dra. Rexarchs Dolores (UAB - España)  
Dra. Reyes Nora (UNSL - Argentina)  
Dra. Tarouco Liane Margarida (UFRGS - Brasil)  
Dra. Russo Claudia (UNNOBA - Argentina)  
Dra. San Martín Patricia (UNR - Argentina)  
Dra. Sanz Cecilia Verónica (UNLP - Argentina)



[www.postgrado.info.unlp.edu.ar/libro-de-tesis-y-tesistas](http://www.postgrado.info.unlp.edu.ar/libro-de-tesis-y-tesistas)

 [www.postgrado.info.unlp.edu.ar](http://www.postgrado.info.unlp.edu.ar)

 [postgrado@lidi.info.unlp.edu.ar](mailto:postgrado@lidi.info.unlp.edu.ar)

 [postinfoUNLP](https://www.instagram.com/postinfoUNLP)

 [postinfoUNLP](https://www.x.com/postinfoUNLP)

 [postgrado-facultad-de-informatica-unlp](https://www.linkedin.com/company/postgrado-facultad-de-informatica-unlp)

 [postgradoinformáticaunlp](https://www.youtube.com/postgradoinformáticaunlp)

 [postgradoinformáticaUNLP](https://www.facebook.com/postgradoinformáticaUNLP)