

2022

# TESIS Y TESISISTAS

DOCTORADO - MAESTRÍAS - ESPECIALIZACIONES



# Índice

## TESIS Y TESISISTAS 2022

---

**Equipo Editorial** pag. 4

---

**Nota Editorial** pag. 5

---

### DOCTORADO EN CIENCIAS INFORMÁTICAS

---

**Dra. María Roxana Martínez** pag. 8

Métricas de calidad para validar los conjuntos de datos abiertos públicos gubernamentales

Quality Metrics to Validate Government Public Open Datasets

---

**Dr. Ariel Cristian Pasini** pag. 10

Modelo de madurez para servicios de gobierno electrónico en el ámbito universitario

E-government maturity model at university

---

**Dr. Jonathan Martín** pag. 12

Apoyo de toma de decisión a comunidades de construcción de conocimiento

Decision-making support to knowledge-building communities

---

**Dra. María José Basgall** pag. 14

Análisis y diseño de técnicas de preprocesamiento de instancias escalables para problemas no balanceados en Big Data. Aplicaciones en situaciones de emergencias humanitarias

Analysis and design of scalable pre-processing techniques of instances for imbalanced Big Data problems. Applications in humanitarian emergencies situations

---

**Dr. Luis Mariano Bibbó** pag. 18

Modelado de Sistemas Colaborativos

Collaborative Systems Modeling

---

**Dr. Giovanni Daián Rottoli** pag. 22

Formulación de Procesos para una Ingeniería de Explotación de Información Espacial

Formulation of Processes for Spatial Knowledge Discovery Engineering

---

**Dr. Nahuel González** pag. 26

Generalización del Modelado de Cadencias de Teclado con Contextos Finitos para su Utilización en Ataques de Presentación y Canal Lateral

Generalizing the Finite Context Modelling Method for Keystroke Dynamics to Presentation and Side-channel Attacks

---

**Dr. Pablo José Iuliano** pag. 30

Cálculo Científico Distribuido sobre Clientes Móviles Indeterminados en redes MANETS

Scientific Calculation Distributed on Undetermined Mobile Clients in MANETS networks

---

**Dra. Marisa Daniela Panizzi** pag. 32

DepProMod: Modelo de Proceso de Despliegue de Sistemas de Software

DepProMod: Deployment Process Model for software systems

---

---

**Dra. Gladys Noemí Kaplan** pag. 36

Proceso de requisitos validado empíricamente

Empirically Validated Requirements Process

---

**Dra. Elba María Bodero Poveda** pag. 38

Modelo de madurez para preservación digital a largo plazo aplicando principios de planificación estratégica

Long-term digital preservation maturity model applying strategic planning principles

---

**Dra. Viviana Elizabeth Cajas Cajas** pag. 42

Interfaces Proactivas Móviles desde Legacies Web: Un Enfoque de Diseño Conductual

Mobile Proactive Interfaces from Web Legacies: A Behavioral Design Approach

---

**Dr. Jorge Geovanni Raura Ruiz** pag. 44

Impacto de las características personales de los programadores en la efectividad de Test-Driven-Development (TDD)

Impact of programmers' human factors on the effectiveness of TestDriven-Development (TDD)

---

**Dra. Mónica del Rocío Romero Pazmiño** pag. 46

Hope Project: Realidad Aumentada para Procesos de Enseñanza Aprendizaje en Niños con TEA

Hope Project: Argument reality for Teaching-learning processes in the asd child population.

---

**Dr. Ignacio Martin Gallardo Urbini** pag. 50

Estrategia de Ciberseguridad distribuida, aplicando el concepto de Operación de Inteligencia

Distributed Cybersecurity Strategy, applying the Intelligence Operation theory

---

**Dr. Marco Remigio Pusda Chulde** pag. 52

Algoritmos para Agricultura de Precisión Utilizando Computación de Alto Rendimiento

Algorithms for Precision Agriculture Using High Performance Computing

---

## MAESTRÍAS

### INGENIERÍA DE SOFTWARE

**Mg. Francisco Gindre** pag. 58

Arquitectura de Software en Wallet de Código Abierto para Privacy Coin en Dispositivos Móviles. El caso de estudio de Zcash

Software Architecture for Open Source Mobile Privacy Coin Mobile Wallet applications. The Zcash Case Study

**Mg. Alejandro Martín Aguirre** pag. 60

Modelos de Predicción Avanzados para el cálculo de Reservas en la Industria Aseguradora

Advanced Prediction Models for Reserves in the Insurance Industry

**Mg. Sergio Hernán Valenzuela Cámara** pag. 64

Detección y Clasificación de Enfermedades en el Tomate Mediante Deep Learning y Computer Vision

Detection and Classification of Diseases in Tomato Using Deep Learning and Computer Vision

### TECNOLOGÍA INFORMÁTICA APLICADA EN EDUCACIÓN

**Mg. Nelba Quintana** pag. 66

Estrategias de utilización de un simulador de entrevistas de trabajo para adultos hipoacúsicos con ayudas auditivas

Strategies for the implementation of a job interview simulator for adults with hearing loss using hearing aids

**Mg. Lucas Romano** pag. 68

Realidad Aumentada en Contextos Educativos y su Relación con el Rendimiento Académico Universitario

Augmented Reality in Educational Contexts and its Relationship with University Academic Performance

**Mg. Romina Itsvan** pag. 72

Deserción estudiantil en la UTN-FRLP. Detección temprana mediante un sistema software de gestión tutorial

Student Desertion from the UTN-FRLP. Early detection using a tutorial management software

### REDES DE DATOS

**Mg. Javier Alejandro Carletto** pag. 76

Optimización de Tráfico en Redes Multiservicios Aplicando Técnicas Heurísticas

Traffic optimization in multiservice networks using heuristic techniques

## ESPECIALIZACIONES

### INGENIERÍA DE SOFTWARE

**Esp. Sonia Raquel Santana** pag. 82

Un análisis de enfoques de Validación de Requerimientos

An Analysis of Requirements Validation Approaches

**Esp. Malqui Fernandez Vidal** pag. 86

La Programación Reactiva: Un nuevo enfoque para trabajar con código asíncrono en la programación web

Reactive Programming: A new approach to working with asynchronous code in web programming

### INTELIGENCIA DE DATOS ORIENTADA A BIG DATA

**Esp. Jonatan Emanuel Martinelli** pag. 88

Clasificación de datos desbalanceados. Su aplicación en la predicción de bajas de beneficiarios de un servicio de salud privado

Classification of imbalanced datasets. Their application to predict the number of beneficiaries leaving a private health insurance entity

**Esp. Patricia Elizabeth Prado** pag. 90

Estudio de Técnicas de Agrupamiento en Procesos de Datos a Gran Escala. Su aplicación en la descripción de Casos de COVID-19 registrados en la República Argentina

Study of Clustering Techniques in Large Scale Data Processing. Its application in the description of COVID-19 Cases registered in the Argentine Republic

### COMPUTACIÓN GRÁFICA, IMÁGENES Y VISIÓN POR COMPUTADORA

**Esp. Lucas Benjamin Cicerchia** pag. 92

Clasificación de cultivos en imágenes Landsat utilizando algoritmos de Active Learning

Crop classification on Landsat imagery using Active Learning algorithms

## TECNOLOGÍA INFORMÁTICA APLICADA EN EDUCACIÓN

**Esp. María Inés Kessler** pag. 96

Las estrategias de formación de usuarios mediadas por tecnología informática en bibliotecas de instituciones de educación superior

User training strategies delivered by means of computer technology in libraries embedded in higher education institutions

**Esp. Daniel Antonio Arellano Sarmiento** pag. 100

Un estado del arte y análisis de experiencias de uso de la robótica educativa en la escuela primaria pública de Chile, Uruguay y Argentina

A state of the art and analysis of experiences of the use of educational robotics in public primary schools in Chile, Uruguay and Argentina

**Esp. María Cecilia Pezzini** pag. 102

Tutores inteligentes en la enseñanza: una revisión y análisis en la educación secundaria

Intelligent Tutoring in Teaching: A Review and Analysis in Secondary Education

### REDES Y SEGURIDAD

**Esp. Fernando Ariel Boettner** pag. 106

Mejoras en la seguridad web del usuario mediante el uso de un proxy local

Improvements in user web security through the use of a local proxy

**Jurados Designados** pag. 109

# 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 2022, 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 2022.

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 32 trabajos que se sintetizan y cuyos contenidos completos se pueden obtener del repositorio institucional de la UNLP (SEDI-CI) 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 32 Tesis, 4 corresponden a docentes-investigadores con cargos en la Facultad de Informática de la UNLP y 28 no tienen relación directa con nuestra Facultad.

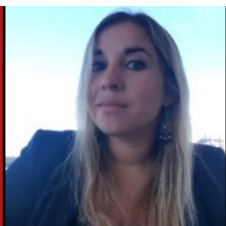
Con este libro esperamos contribuir a la difusión de las actividades académicas de Postgrado, así como a la trasmisión de conocimientos generados en la Facultad de Informática de la UNLP.

## **Equipo Editorial del Postgrado**



01

DOCTORADO EN  
**CIENCIAS INFORMÁTICAS**



**Directora**  
Dra. Rocío Rodríguez

**Codirectora**  
Dra. Claudia Pons

**Asesor Científico**  
Dr. Pablo Vera

**Fecha de aprobación**  
29 de marzo de 2022

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

## Métricas de calidad para validar los conjuntos de datos abiertos públicos gubernamentales

**Palabras clave:** Datos Abiertos; Datos Públicos; Gobierno Abierto; Métricas de Calidad de Datos

### Motivación

El objetivo principal de la presente tesis es proponer una serie de métricas de calidad para validar los conjuntos de datos abiertos públicos gubernamentales.

### Aportes de la tesis

Las contribuciones principales de este trabajo de tesis son las siguientes:

- Relevamiento del estado situación actual de los aspectos más relevantes en el tratamiento de datasets públicos abiertos a nivel nacional como internacional.
- Propuesta de un conjunto de métricas críticas y no críticas para analizar la calidad de datos abiertos;
- Desarrollo propio de una herramienta para validar desde varios aspectos estándares, la calidad de los datasets publicados en portales de datos abiertos estatales;
- Detección de falencias en los datasets gubernamentales disponibilizados;
- Aporte sobre las mejoras en la calidad del dato y concientización de su importancia para una correcta divulgación del contenido público tanto a nivel nacional como internacional.

### Líneas de I/D futuras

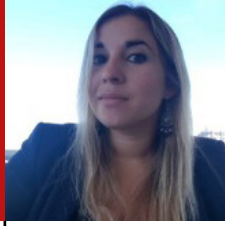
Algunos de los aspectos que pueden ser estudiados como parte de los trabajos futuros de la presente tesis son los siguientes:

- Continuar trabajando en la ampliación de la herramienta de validación HEVDA, incorporando más criterios y princi-

pios mencionados en este trabajo por las distintas organizaciones mundiales que inculcan la cultura de gobierno abierto.

- Incorporar a la herramienta desarrollada más tipos de formatos abiertos, como ser: XML, JSON, etc. lo que llevará a establecer nuevos aspectos de control en las estructuras de los formatos que serán implementados.
- Analizar las opciones de aplicaciones gráficas que podrían ser utilizadas con los datasets para ser embebidos en el código fuente de la herramienta HEVDA.
- Incorporar una herramienta de gráficos estadísticos con el análisis de las métricas críticas y no críticas.
- Analizar los datasets orientados a la geolocalización para el tipo de contexto:  
Coordenadas de longitud y latitud, formatos geoespaciales, archivos del tipo WKT (puntos de coordenadas), SHP (datos geográficos), etc.
- Efectuar un estudio detallado de los datasets geoespaciales para proponer métricas de calidad de datos para este entorno de trabajo.





english

**Advisor**

Dra. Rocío Rodríguez

**Codirector**

Dra. Claudia Pons

**Scientist Consultant**

Dr. Pablo Vera

**Thesis approval date**

March 29, 2022

**SEDICI**

[Full Text](#)

## Quality Metrics to Validate Government Public Open Datasets

**Keywords:** Open Data; Public Data; Open Government; Data Quality Metrics

### Motivation

The main objective of this thesis is to propose a series of quality metrics to validate government public open datasets.

### Thesis contributions

The main contributions of this thesis work are the following:

- Analysis of the current state of the most relevant aspects in the treatment of open public datasets at a national and international level.
- Proposal of a set of critical and non-critical metrics to analyze the quality of open data;
- Own development of a tool to validate from several standard aspects, the quality of the datasets published in state open data portals;
- Detection of shortcomings in the available government datasets;
- Contribution on improvements in the quality of the data and awareness of its importance for a correct disclosure of public content both nationally and internationally.

### Future Research Lines

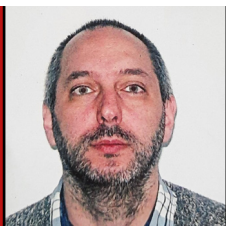
Some of the aspects that can be studied as part of the future work of this thesis are the following:

- Continue working on the expansion of the HEVDA validation tool, incorporating more criteria and principles mentioned in this work by the different world organizations that instill the culture of open government.
- Incorporate more types of open formats into the developed

tool, such as: XML, JSON, etc. which will lead to establishing new aspects of control in the structures of the formats that will be implemented.

- Analyze the graphical application options that could be used with the datasets to be embedded in the source code of the HEVDA tool.
- Incorporate a statistical graph tool with the analysis of critical and non-critical metrics.
- Analyze geolocation-oriented datasets for the type of context: Longitude and latitude coordinates, geospatial formats, files such as WKT (coordinate points), SHP (geographical data), etc.
- Carry out a detailed study of the geospatial datasets to propose data quality metrics for this work environment.

## DOCTORADO EN CIENCIAS INFORMÁTICAS



**Dr. Ariel Cristian Pasini**

e-mail

apasini@lidi.info.unlp.edu.ar

español

**Directoras**  
Dra. Elsa Estevez  
Lic. Patricia Pesado

**Fecha de aprobación**  
30 de marzo de 2022

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

## Modelo de madurez para servicios de gobierno electrónico en el ámbito universitario

**Palabras clave:** Gobierno electrónico; servicios públicos; gobierno universitario; servicios universitarios; modelo de madurez de gobierno electrónico; madurez de servicios universitarios

### Motivación

Con la evolución de las TICs, el concepto de gobierno electrónico ha ido evolucionando hasta transformarse en lo que hoy se denomina gobernanza digital contextualizada, o dirigida por políticas. Generalmente se utiliza el término “gobierno” en la estructura gubernamental de un Estado (municipios, provincias, países), pero también existen gobiernos dentro de diferentes contextos, como agencias impositivas, agencias de control gubernamental, agencias de administración de justicia, etc., que tienen cierto grado de autonomía del Estado. Dentro de este tipo de estructuras se encuentran las Universidades Nacionales, que poseen su propio gobierno democrático interno y gozan de autonomía del gobierno político del Estado. El gobierno universitario está compuesto por alumnos, docentes, no docentes, graduados, alumnos de posgrado, etc., dependiendo de la estructura definida en su estatuto. Todos en su conjunto representan a los ciudadanos que desarrollan sus actividades en el marco de las reglamentaciones que dispone dicho gobierno. Para llevar a cabo el cumplimiento de las reglamentaciones que exige el gobierno universitario, la Universidad pone a disposición de cada tipo de ciudadano un gran número de servicios. Hoy en día, estos servicios son brindados a través del uso de las TICs, lo que conlleva a clasificar a estas prestaciones como servicios de gobierno electrónico universitario. Tratándose de una comunidad reducida, nos permite hacer evaluación de la calidad de las prestaciones de estos servicios de forma más certera. El objetivo principal de esta tesis es definir un modelo de evaluación, basado en los aplicados en estructuras de gobiernos masivas, que permita clasificar en niveles de madurez a los servicios prestados por las unidades académicas y, en consecuencia, permita definir un escalafón de unidades académicas en la prestación de servicios con el fin de ofrecer recomendaciones para mejorar la prestación de servicios a su comunidad.

### Aportes de la tesis

Las principales contribuciones de la tesis son la vinculación de los conceptos de gobierno electrónico al ámbito del contexto de los gobiernos universitarios y la propuesta del modelo de evaluación de madurez para servicios de gobierno electrónico en el ámbito universitario.

### Líneas de I/D futuras

Como trabajos futuros se plantean tres líneas principales:

- 1.** Ampliación del modelo básico a un conjunto de servicios que involucren más representantes de la comunidad universitaria. En casos particulares se detectaron actividades que eran relevantes para las unidades académicas que no pudieron ser evaluadas por no estar dentro de los servicios básicos contemplados en el modelo.
- 2.** Generar herramientas de asistencia para unidades académicas que faciliten la migración de servicios, que actualmente se realizan por mostrador, a servicios electrónicos universitarios. Varios de los servicios que se realizan por mostrador se repetían en diferentes unidades académicas. Podría ser de utilidad generar procesos reutilizables que le faciliten a la unidad académica pasar de prestar el servicio por mostrador a brindarlo como un servicio electrónico.
- 3.** Generar herramientas para que la unidad académica pueda realizar autoevaluaciones en la prestación de sus servicios electrónicos universitarios. Que la unidad académica pueda gestionar sus propias mejoras, registrando el crecimiento en la madurez de sus servicios universitarios, puede generar un importante incentivo para permanecer en el proceso de mejora continua.



**Advisors**  
Dra. Elsa Estevez  
Lic. Patricia Pesado

**Thesis approval date**  
March 30, 2022

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

## **E-government maturity model at university**

**Keywords:** E-government; university government; publics services, university services; maturity model

### **Motivation**

With the evolution of ICTs, the concept of electronic government has evolved into what is now called contextualized or policy-driven digital governance. The term “government” is generally used in the governmental structure of a State (municipalities, provinces, countries), but there are also governments within different contexts, such as tax agencies, government control agencies, justice administration agencies, etc., that have a certain autonomy from the state. Within this type of structures are the National Universities. They have their own internal democratic government and they enjoy autonomy from the political government of the State. The university government is made up of students, teachers, non-teachers, graduates, postgraduate students, etc., depending on the structure defined in its statute, and all together represent the citizens who carry out their activities within the framework of the regulations that the government has. To carry out compliance with the regulations required by the university government, the University makes a large number of services available to each type of citizen. Today, these services are provided through the use of ICTs, which leads to classify these benefits as university electronic government services. Being a small community, it allows us to evaluate the quality of the services provided more accurately.

The main objective of this thesis is to define an evaluation model, based on those applied in massive government structures, that allows classifying the services provided by academic units into maturity levels and, consequently, allows defining a scale of academic units in the service delivery in order to offer recommendations that allow them to improve the delivery of services to their community

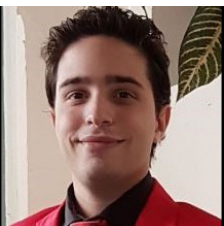
### **Thesis contributions**

The main contributions of the thesis are the linking of e-government concepts to the context of university governments and the proposal of the maturity evaluation model for e-government services in the university environment.

### **Future Research Lines**

As future works, three main lines are proposed:

- 1.** Expansion of the basic model to a set of services that involve more representatives of the university community. In particular cases, activities were detected that were relevant for the academic units that could not be evaluated because they were not within the basic services contemplated in the model.
- 2.** Generate assistance tools for academic units that facilitate the migration of services, which are currently performed in person, to university electronic services. Several of the services that are performed in person were repeated in different academic units. It could be useful to generate reusable processes that make it easier for the academic unit to go from providing the service in person to providing it as an electronic service.
- 3.** Generate tools so that the academic unit can perform self-evaluations in the provision of its university electronic services. That the academic unit can manage its own improvements, registering the growth in the maturity of its university services, can generate an important incentive to stay in the continuous improvement process.



**Directores**  
Dr. Diego Torres  
Dr. Alejandro Fernández

**Fecha de aprobación**  
21 de abril de 2022

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

## Apoyo de toma de decisión a comunidades de construcción de conocimiento

**Palabras clave:** Aprendizaje por refuerzo profundo; Ludificación; Comunidades de construcción de conocimiento; Optimización; Stack Overflow; Wikipedia

### Motivación

En este trabajo, se busca apoyar la toma de decisión a fin de lograr una “ludificación-óptima” desde la base de una estrategia de ludificación preexistente y tomando en consideración el contexto de incertidumbre inherente a la forma de trabajo colaborativo en las Comunidades de construcción de conocimiento (CCC).

Por lo tanto, este trabajo busca comprobar a partir de aplicar agentes de aprendizaje por refuerzo en un entorno que simule de una estrategia de ludificación, en base a datos históricos, se puede apoyar la selección de variables de configuración óptimas para integrar la estrategia de ludificación en una CCC manteniendo su comportamiento intrínseco.

De esto se desprenden las siguientes afirmaciones a validar como subobjetivos:

- Existen patrones de comportamiento de los miembros de las CCC los cuales pueden encontrarse en sus datos históricos de actividad.
- A partir de los datos históricos del comportamiento de la comunidad también se podrá captar y mantener comportamientos relacionados a una estrategia de ludificación distinta a la cual se pretende configurar. Realizando una transferencia de conceptos de ludificación de una estrategia a otra.
- El enfoque de solución introducido en este trabajo es flexible a la integración de nuevas estrategias y aplicación de futuros avances en el área de aprendizaje por refuerzo.

### Aportes de la tesis

La configuración de la tabla de mapeo entre acciones de la CCC objetivo y las acciones de la estrategia de ludificación a simular presentan la problemática de que actualmente se realizan de forma manual. Esto genera la necesidad de poseer conocimiento sobre ambas partes y la posible relación entre ellas. Sería de interés poder evitar esta necesidad de conocimiento, así como el proceso manual de crear ese mapeo de información.

Además, el framework fue desarrollado como tal con el objetivo de poder ser extendido en múltiples aspectos. Por ejemplo, desarrollando un compendio de herramientas generalizadoras que faciliten la normalización de cualquier estado o recompensa creada con el framework.

También en esta tesis se trabajó sobre una única simulación de una estrategia de ludificación basada en Metagame. Por lo cual, queda abierta la problemática de desarrollar otras simulaciones para estrategias de ludificación preexistentes o nuevas, que sean más adaptables o configurables, sobre las cuales evaluar también el enfoque y el framework.

Finalmente, basándose en el enfoque, pero por fuera de los objetivos de esta tesis. Sería interesante analizar este enfoque de recomendación de configuraciones mediante aprendizaje por refuerzo, basándose en simulaciones con datos históricos, sobre la problemática de configurar aplicaciones que no sean pertenecientes al área de ludificación.

### Líneas de I/D futuras

Algunas de las posibles líneas futuras que surgen de este trabajo son:

- Automatizar el proceso manual de mapear de acciones entre las comunidades y la simulación.
- Extender el framework con nuevas recompensas y formatos de estados.
- Además, esto recae también en encontrar cuál de estas recompensas y estados son óptimas llevándonos a un espacio de meta-configuración.
- Evaluar el enfoque con nuevas ludificaciones.
- Finalmente, si bien este enfoque ataca la problemática de la configuración de ludificaciones, el mismo es factible de implementar para la configuración automática de otros tipos de sistemas.



english

**Advisors**  
Dr. Diego Torres  
Dr. Alejandro Fernández

**Thesis approval date**  
April 21, 2022

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

## Decision-making support to knowledge-building communities

**Keywords:** Deep reinforcement learning; Gamification; Knowledge building community; Optimization; Stack Overflow; Wikipedia

### Motivation

In this thesis, we seek to support decision making in order to achieve “optimal gamification” on the basis of a pre-existing gamification strategy and taking into consideration the context of uncertainty inherent to the collaborative way of working in Knowledge Construction Communities (KCCs).

Therefore, this work seeks to verify that by applying reinforcement learning agents in an environment that simulates a gamification strategy, based on historical data, it is possible to support the selection of optimal configuration variables to integrate the gamification strategy in a KCC while maintaining its intrinsic behaviour.

From this, the following statements can be validated as sub-objectives:

- There are behavioural patterns of KCC members which can be found in their historical activity data.
- From the historical data of the community’s behaviour, it will also be possible to capture and maintain behaviours related to a different gamification strategy to the one that is intended to be configured. By transferring gamification concepts from one strategy to another.
- The solution approach introduced in this work is flexible to the integration of new strategies and application of future developments in the area of reinforcement learning.

### Thesis contributions

This thesis work presents a methodology to provide decision support in the context of inserting a gamification strategy in a KCC. As part of this methodology, simulated applications of a gamification strategy in different communities are presented. A framework is also developed to facilitate the implementation of the proposed methodology. This framework allowed to study the performance of the proposed methodology, providing flexibility and extensibility for the application of the methodology in different KCCs.

An evaluation of this framework in the Wikipedia KCC is

presented. During this evaluation, the successful generation of configurations by the trained agent in the different subcommunities is verified. To validate the generalisability to other KCCs, an evaluation of the framework in the Stack Overflow KCC is presented. The effectiveness of the configuration found for subsequent periods of time in the same community is evaluated. This evaluation shows that the obtained configurations are effective in later time periods in the same community.

Finally, the extensibility of the approach and the framework in terms of new strategies or approaches in the area of reinforcement learning is validated. This evaluation demonstrates the ease of integration of new agents regardless of the technology that supports them.

### Future Research Lines

Some of the possible future directions that emerge from this work are:

- Automate the manual process of mapping actions between communities and the simulation.
- Extend the framework with new rewards and state formats.
- Furthermore, this also relies on finding which of these rewards and states are optimal leading to a meta-configuration space.
- Evaluate the approach with new gamification.
- Finally, while this approach addresses the problem of gamification configuration, it is feasible to implement it for the automatic configuration of other types of systems.



**Directores**  
Dr. Marcelo Naiouf  
Dr. Alberto Fernández Hilario

**Fecha de aprobación**  
5 de mayo de 2022

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

## **Análisis y diseño de técnicas de preprocesamiento de instancias escalables para problemas no balanceados en Big Data. Aplicaciones en situaciones de emergencias humanitarias**

**Palabras clave:** Big Data; Machine Learning; Data Preprocessing; Data Imbalance; Data reduction; Scalability

### **Motivación**

El enorme volumen de datos provenientes de distintas fuentes, realmente variado en su tipología, y que se genera y procesa a gran velocidad, es lo que se conoce como Big Data. La importancia de los datos está en extraer conocimiento de ellos, en aras de facilitar la toma de decisiones. Para ello, es imprescindible la aplicación de modelos de Machine Learning. Sin embargo, hay que tener en cuenta que las técnicas que hasta ahora ofrecían buenos resultados, no siempre soportan manejar Big Data por temas de escalabilidad. Es por ello que necesitan ser adaptadas para trabajar en entornos distribuidos, o se deben crear nuevas técnicas o estrategias para lidiar con este nuevo escenario. Asimismo, los conjuntos de datos normalmente pueden tener ciertas características o complejidades no deseadas que interfieren en la efectividad del proceso de extracción del conocimiento, y es por ello que deben ser preprocesados debido a que la mayoría de los modelos de aprendizaje asumen que los datos están libres de esas características. Por lo anterior, y dado que existen pocas soluciones escalables y capaces de manejar Big Data relacionadas a este tema, esta tesis aborda al preprocesamiento distribuido y escalable de conjuntos Big Data, con el fin de obtener datos de buena calidad, conocido como Smart Data. Particularmente se centra en los problemas de clasificación, y en atender a las siguientes características: (a) datos no balanceados; (b) redundancia; (c) alta dimensionalidad; y (d) solapamiento.

Para ello, se establecen los siguientes objetivos específicos: Habilitar que un algoritmo del estado del arte altamente empleado para el tratamiento de la distribución de clases en escenarios de datos tradicionales (Small Data), sea

capaz de obtener resultados adecuados a partir de grandes conjuntos de datos de manera distribuida y en tiempos de ejecución razonables.

Diseñar e implementar una metodología rápida y escalable para la reducción tanto en instancias como en atributos para los conjuntos Big Data que presentan alta redundancia y dimensionalidad, manteniendo la capacidad predictiva del conjunto de datos original.

Diseñar e implementar una estrategia para la caracterización escalable de los datos en el contexto de clasificación en Big Data, con foco en las zonas ambiguas del problema.

Aplicar el conocimiento adquirido durante la fase de desarrollo para resolver problemas de interés relacionados con emergencias humanitarias.

### **Aportes de la tesis**

Respecto a los datos no balanceados, se presentó SMOTE-BD, un SMOTE para Big Data basado en un estudio sobre las particularidades necesarias para que su diseño sea totalmente escalable, y que además su comportamiento se ajuste lo más fielmente posible a la técnica secuencial del estado del arte (tan popular en escenarios Small Data). Asimismo, se introdujo una variante de SMOTE-BD, denominada SMOTE-MR, que sigue un diseño tal que procesa los datos localmente en cada nodo.

En relación a la redundancia y a la alta dimensionalidad, se presentó FDR2-BD, una metodología escalable para reducir un conjunto Big Data de manera dual (reducción de atributos y de instancias), con la premisa de mantener la calidad predictiva respecto de los datos originales. La propuesta se basa en un esquema de validación cruzada



donde se realiza un proceso de hiperparametrización. FDR2-BD permite conocer si un conjunto de datos dado es reducible manteniendo el poder predictivo de los datos originales dentro de un umbral. Por consiguiente, nuestra propuesta informa cuáles son los atributos de los datos de mayor importancia y cuál es el porcentaje de reducción uniforme de instancias que se puede llevar a cabo.

Los resultados mostraron la fortaleza de FDR2-BD obteniendo valores de reducción muy elevados para la mayoría de los conjuntos de datos estudiados, tanto en lo que respecta a la dimensionalidad como a los porcentajes de reducción de instancias propuestos. En concreto, se alcanzó alrededor del 70 % de reducción de las características y 98 % de reducción de las instancias, para un umbral de pérdida predictiva máxima aceptada del 1 %.

En cuanto al solapamiento, se presentó GridOverlap-BD, una metodología para la caracterización escalable de problemas Big Data de clasificación, que se apoya en el particionamiento del espacio de características basado en rejilla. GridOverlap-BD permite identificar o caracterizar las áreas del problema en dos tipologías: puras y solapadas. Además, se introdujo una métrica de complejidad derivada de aplicar GridOverlap-BD, con foco en cuantificar el solapamiento presente en los datos.

De la experimentación se observó que tanto la caracterización de las zonas de un problema como la cuantificación del grado de solapamiento se llevaron a cabo de manera efectiva para los conjuntos de datos utilizados. Ello implica una aproximación pionera escalable y totalmente agnóstica (independiente del modelo) para la caracterización de las instancias de un problema Big Data.

## Líneas de I/D futuras

Puesto que hay pocas soluciones para el preprocesamiento de datos en Big Data, y menos aún para problemas de clasificación con múltiples clases, nos planteamos como trabajo a futuro extender todas nuestras propuestas para problemas de este tipo.

En relación al equilibrio de clases, se podría llevar a cabo el desarrollo de modelos híbridos (combinación entre sobre y bajo-muestreo) centrados en las zonas donde el remuestreo es especialmente necesario. Asimismo, combinar la tarea de remuestreo con un paso previo de selección de las características más relevantes del problema para abarcar en conjunto a los datos no balanceados y la alta dimensionalidad normalmente presente en los conjuntos Big Data.

En cuanto a la reducción de los datos manteniendo su poder predictivo, se propone aplicar sobre el conjunto reducido la identificación de áreas solapadas o alguna otra caracterización de los datos a definir, y así intentar mejorar aún más la calidad de los datos.

Con respecto a GridOverlap-BD, se podría desagregar los tipos actuales de celdas en más subtipos, considerando distintos escenarios de proporciones entre clases. A su vez, se podrían emplear otras métricas de complejidad de los datos como guías para seleccionar una granularidad que se adecúe al conjunto Big Data con el que se esté experimentando. Además, se propone el diseño de un clasificador ensemble a partir de la identificación de clase de una instancia dada a través de distintas granularidades, para su posterior agregación y obtención de etiqueta de clase final. Este tipo de solución puede ser útil no solo con el objetivo de mejorar la construcción del modelo, sino incluso para identificar ejemplos de etiqueta dudosa que puedan ser causa de algún tipo de sesgo o falta de equidad en los datos.



**Advisors**  
Dr. Marcelo Naiouf  
Dr. Alberto Fernández Hilario

**Thesis approval date**  
May 5, 2022

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

## **Analysis and design of scalable pre-processing techniques of instances for imbalanced Big Data problems. Applications in humanitarian emergencies situations**

**Keywords:** Big Data; Machine Learning; Data Preprocessing; Data Imbalance; Data reduction; Scalability

### **Motivation**

The enormous volume of data from different sources, really varied in its typology, and generated and processed at great speed, is known as Big Data. The importance of data lies in extracting knowledge from it, in order to facilitate decision-making. To do this, applying Machine Learning models is essential. However, it must be noted that the techniques that until now have offered good results are not always able to handle Big Data due to scalability issues. For this reason, they need to be adapted to work in distributed environments, or new techniques or strategies need to be created to deal with this new scenario.

In addition, datasets can usually have certain undesired characteristics or complexities that interfere with the effectiveness of the knowledge extraction process, so they must be preprocessed due to the fact that most learning models assume that the data are free of those characteristics.

Therefore, and since there are few scalable solutions capable of handling Big Data related to this topic, this thesis addresses the distributed and scalable pre-processing of Big Data sets, in order to obtain good quality data, known as Smart Data. Particularly, it focuses on classification problems, and on addressing the following characteristics: (a) imbalanced data; (b) redundancy; (c) high dimensionality; and (d) overlapping.

The following specific objectives are established for this purpose: Enable a state-of-the-art algorithm widely used for the treatment of class distribution in traditional data scenarios (Small Data), to be able to obtain adequate results from large datasets in a distributed manner and in reasonable execution times.

To design and to implement a fast and scalable methodology for the reduction in both instances and attributes for Big Data sets with high redundancy and dimensionality, while

maintaining the predictive capacity of the original dataset. To design and to implement a strategy for scalable data characterisation in the context of Big Data classification, focusing on the ambiguous areas of the problem.

To apply the knowledge acquired during the development phase to solve problems of interest related to humanitarian emergencies.

### **Thesis contributions**

Regarding imbalanced data, SMOTE-BD was presented, a SMOTE for Big Data based on a study of the necessary particularities for its design to be fully scalable, and also its behaviour to match as closely as possible the sequential technique of the state of the art (so popular in Small Data scenarios). A variant of SMOTE-BD, called SMOTE-MR, was also introduced, which follows a design that processes data locally at each node.

In relation to redundancy and high dimensionality, FDR2-BD was presented, a scalable methodology to reduce a Big Data set in a dual way (reduction of attributes and instances), with the premise of maintaining the predictive quality with respect to the original data. The proposal is based on a cross-validation scheme where a hyperparameterisation process is conducted. FDR2-BD allows us to know if a given dataset is reducible while maintaining the predictive power of the original data within a threshold. Therefore, our proposal informs which data attributes are the most important ones and what is the percentage of uniform instance reduction that can be performed.

The results showed the strength of FDR2-BD by obtaining very high reduction values for most of the studied datasets, both



in terms of dimensionality and proposed instance reduction percentages. In particular, around 70 % feature reduction and 98 % instance reduction were achieved, for a maximum accepted predictive loss threshold of 1 %.

Regarding overlapping, GridOverlap-BD was presented, a methodology for scalable characterisation of Big Data classification problems, which relies on grid-based feature space partitioning. GridOverlap-BD allows to identify or characterise problem areas in two typologies: pure and overlapping. In addition, a complexity metric derived from applying GridOverlap-BD was introduced, with focus on quantifying the overlap present in the data.

From the experimentation, it was observed that both the characterisation of problem areas and the quantification of the degree of overlap were effectively performed for the datasets used. This implies a pioneering scalable and fully agnostic (model-independent) approach for the characterisation of Big Data problem instances.

## Future Research Lines

Since there are few solutions for data pre-processing in Big Data, and even fewer for multi-class classification problems, we envisage as future work to extend all our proposals to problems of this type.

In relation to class balancing, the development of hybrid models (combination between oversampling and undersampling) could be performed focusing on the areas where resampling is especially necessary. Also, combining the resampling task with a previous step of selecting the most relevant features of the problem to encompass together the imbalanced data and the high dimensionality normally present in Big Data sets.

Regarding the reduction of the data while maintaining its predictive power, it is proposed to apply on the reduced set the identification of overlapping areas or some other characterisation of the data to be defined, and thus try to further improve the quality of the data.

With respect to GridOverlap-BD, the current cell types could be disaggregated into more subtypes, considering different scenarios of proportions between classes. At the same time, other data complexity metrics could be used as guides to select a granularity that suits the Big Data set being experimented with. In addition, we propose the design of an ensemble classifier based on the class identification of a given instance through different granularities, for its subsequent aggregation and obtaining a final class label. This type of solution can be useful not only for the purpose of improving model building, but even for identifying examples of dubious labeling that may be the cause of some kind of bias or unfairness in the data.



**Directoras**  
Dra. Claudia Pons  
Dra. Roxana Giandini

**Codirector**  
Dr. Gustavo Héctor Rossi

**Fecha de aprobación**  
10 de mayo de 2022

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

## Modelado de Sistemas Colaborativos

**Palabras clave:** Ingeniería de Software, administración de Proyectos en TI, desarrollo Ágil, desarrollo dirigido por modelos, estimación de proyectos

### Motivación

El objetivo principal de la presente tesis es contribuir con recursos de ingeniería de software que guíen la construcción de sistemas colaborativos complejos y que faciliten su desarrollo.

Para llevar a cabo dicho objetivo se plantean los siguientes objetivos específicos:

- Estudiar los sistemas colaborativos, los conceptos centrales y los modelos conceptuales que los describen;
- Relevar y analizar los trabajos relacionados en la temática de Sistemas Colaborativos;
- Estudiar metodologías y paradigmas apropiados para describir los Sistemas Colaborativos;
- Identificar mecanismos que describan los Sistemas Colaborativos de manera precisa, concisa y amigable;
- Proponer un lenguaje específico (DSL) y notaciones formales para describir distintas componentes de los Sistemas Colaborativos;
- Desarrollar herramientas para diseñar los Sistemas Colaborativos en el marco del desarrollo dirigido por modelos;
- Proponer métricas de evaluación, validación y verificación cuantitativas y cualitativas del lenguaje propuesto;
- Estudiar tecnologías para plataformas específicas que puedan soportar funcionalidad típica de los Sistemas Colaborativos;

- Modelizar las transformaciones de los conceptos del lenguaje específico hacia posibles tecnologías de implementación;
- Desarrollar una semántica que interprete los modelos producidos por el lenguaje propuesto y obtenga otros modelos derivados;
- Analizar los resultados arrojados con las transformaciones ejecutadas.

### Aportes de la tesis

Relevamiento del estado situación actual del modelado de Sistemas Colaborativos.

- Proponer un lenguaje específico de dominio (CSSL V2.0) que describa a los Sistemas Colaborativos, especialmente aspectos dinámicos como procesos colaborativos, protocolos y awareness;
- Desarrollar un conjunto de editores gráficos como una sintaxis concreta asociada a CSSL, que permite construir modelos de Sistemas Colaborativos;
- Desarrollar un esquema de transformación de los conceptos del lenguaje a conceptos de una plataforma y tecnología específica, como se define en MDD.
- Desarrollar una semántica que interprete los modelos escritos con CSSL y obtenga versiones Web ejecutables para una plataforma particular;

- Obtener métricas que evalúen, validen y verifique el lenguaje desarrollado;
- Aportar sobre las mejoras en la producción de software utilizando el desarrollo dirigido por modelos

### **Líneas de I/D futuras**

- Ampliar el relevamiento y estudio de nuevos aspectos que surjan en base a las nuevas necesidades de los Sistemas Colaborativos, manteniendo el lenguaje CSSL actualizado.
- Continuar trabajando en la ampliación de los editores gráficos que utilizan el lenguaje.
- Incorporar y vincular otros perfiles de UML que complementen el diseño de Sistemas Colaborativos, por ejemplo perfiles para aplicaciones Mobile.
- Desarrollar otras semánticas del lenguaje que permita obtener versiones ejecutables en otras plataformas y utilizando otras tecnologías.
- Incorporar herramientas de evaluación de usabilidad, especialmente para evaluar el impacto de la funcionalidad de awareness en distintos dominios (educación, juegos colaborativos, etc.)
- Incorporar funcionalidad colaborativa a los editores gráficos que trabajan con el lenguaje CSSL.
- Desarrollar un diseño de una arquitectura para soportar servicios de awareness para sistemas colaborativos de gran escala.



**Advisors**  
Dra. Claudia Pons  
Dra. Roxana Giandini

**Codirector**  
Dr. Gustavo Héctor Rossi

**Thesis approval date**  
May 10, 2022

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

## Collaborative Systems Modeling

**Keywords:** Software Engineering, IT Project Management, Agile Development, Model Driven Development, Project Estimation

### Motivation

The main objective of this thesis is to contribute software engineering resources that guide the construction of complex collaborative systems and facilitate their development.

To carry out this objective, the following specific objectives are proposed:

- Study collaborative systems, central concepts and conceptual models that describe them;
- Survey and analyze related works on the subject of Collaborative Systems;
- Study appropriate methodologies and paradigms to describe Collaborative Systems;
- Identify mechanisms that describe Collaborative Systems in a precise, concise and friendly way;
- Propose a specific language (DSL) and formal notations to describe different components of Collaborative Systems;
- Develop tools to design Collaborative Systems within the framework of model-driven development;
- Propose quantitative and qualitative evaluation, validation and verification metrics of the proposed language;
- Study technologies for specific platforms that can support typical functionality of Collaborative Systems;
- Model the transformations of the specific language concepts towards possible implementation technologies;

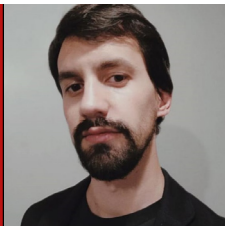
- Develop semantics that interpret the models produced by the proposed language and obtain other derived models;
- Analyze the results obtained with the transformations carried out.

### Thesis contributions

- Survey of the current situation of modeling of Collaborative Systems;
- Propose a domain specific language (CSSL V2.0) that describes Collaborative Systems, especially dynamic aspects such as collaborative processes, protocols and awareness;
- Develop a set of graphic editors as a specific syntax associated with CSSL, which allows to build models of Collaborative Systems;
- Develop a scheme for the transformation of the language concepts to concepts of a specific platform and technology, as defined in MDD.
- Develop semantics that interpret models written with CSSL and obtain executable Web versions for a particular platform;
- Obtain metrics that evaluate, validate and verify the language developed;
- Contribute on improvements in software production using model-driven development.

### **Future Research Lines**

- Expand the survey and study of new aspects that arise based on the new needs of Collaborative Systems, keeping the CSSL language updated.
- Continue working on expanding the graphic editors that use the language.
- Incorporate and link other UML profiles that complement the design of Collaborative Systems, for example profiles for Mobile applications.
- Develop other semantics of the language that allow us to obtain executable versions on other platforms and using other technologies.
- Incorporate usability evaluation tools, especially to evaluate the impact of awareness functionality in different domains (education, collaborative games, etc.).
- Incorporate collaborative functionality to graphic editors that work with the CSSL language.
- Develop an architecture design to support awareness services for large-scale collaborative systems.



**Director**  
Mg. Rodolfo Bertone

**Asesor Científico**  
Dr. Hernán Merlino

**Fecha de aprobación**  
17 de mayo de 2022

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

## Formulación de Procesos para una Ingeniería de Explotación de Información Espacial

**Palabras clave:** Explotación de Información; Procesos de Explotación de Información; Datos Espaciales; Datos geolocalizados; Minería de Datos Espaciales

### Motivación

Los procesos de explotación de información, estructuran de una forma lógica las actividades a ser llevadas a cabo sobre los conjuntos de datos para la extracción de conocimiento no trivial e implícito en los datos presentes en los repositorios de las organizaciones.

Los datos espaciales son datos dependientes del contexto que modelan eventos, objetos o situaciones localizados en un espacio determinado. Usándolos es posible extraer patrones que involucren tanto a sus atributos como también a las relaciones espaciales entre ellos, pero aumentando a su vez la complejidad del proceso.

Si bien existen algoritmos que permiten extraer estas relaciones complejas, usualmente se ignoran múltiples aspectos conceptuales de valor. Por lo general, se aplican algoritmos de minería de datos tradicionales siguiendo procesos estructurados no pensados para manejar las peculiaridades asociadas a este tipo de información.

Se plantea como objetivo desarrollar y validar Procesos de Explotación de Información Espacial que sistematicen la obtención de conocimiento sin ignorar las propiedades particulares de este tipo de datos y, específicamente:

- Identificar los tipos de patrones que se pueden obtener a partir de este tipo de datos.
- Identificar las propiedades de los datos espaciales que aportan valor para la toma de decisiones estratégicas.
- Diseñar procesos de explotación de información para la obtención de patrones de conocimiento a partir de datos espaciales haciendo uso de técnicas y soluciones implementadas en las herramientas tecnológicas actuales.

### Aportes de la tesis

Se han presentado cuatro procesos de explotación de información espacialmente referenciada.

- Un proceso para el descubrimiento de grupos espaciales que permite distinguir entre grupos, regiones y zonas calientes y luego describir cada grupo usando los atributos no espaciales de los datos.
- Un proceso para el descubrimiento de anomalías espaciales que posibilita la búsqueda de anomalías locales, la descripción del comportamiento de los datos normales y la búsqueda de grupos de datos espaciales con comportamiento anómalo.
- Un proceso para el descubrimiento de asociaciones espaciales locales, la consideración de múltiples tipos de relaciones espaciales simultáneas y la posibilidad de utilizar descripciones difusas de las mismas.
- Un proceso para el descubrimiento de patrones de co-localización en contextos locales extendible a otros tipos de relaciones.

Todos estos artefactos utilizan datos vectoriales como entrada. Además, consideran la posibilidad de representar múltiples relaciones espaciales, generalmente haciendo uso de teoría de grafos. Estas relaciones a su vez pueden tener naturaleza rígida o difusa. También permiten tener en consideración el fenómeno de la autocorrelación espacial y el fenómeno de heterogeneidad espacial. Por último, los procesos son independientes de algoritmos y tecnologías particulares.

### Líneas de I/D futuras

En el ciclo de diseño de estos procesos han surgido las siguientes

inquietudes que no han sido consideradas en el alcance de este trabajo y han quedado pendientes para trabajos futuros:

- Existe una amplia variedad de patrones que no han sido consideradas. En particular, resulta necesario abordar la clasificación de datos espacialmente referenciados.
- Es necesario un conjunto de procesos que opere con datos de entrada de tipo raster, sin necesidad de conversión entre tipos de datos.
- Los artefactos diseñados son generales para adaptarse a una gran cantidad de situaciones particulares, pero para ciertos problemas es necesario utilizar abordajes específicos. Se requiere un estudio sistemático de la bibliografía que mapee estos problemas particulares con los algoritmos que han sido propuestos para solucionarlos en el estado del arte.
- Existen otros fenómenos o características espaciales que no han sido exploradas y que ante ciertos escenarios pueden resultar de interés.
- Se requiere, por último, una revisión periódica de estos procesos para su actualización, en virtud de las nuevas tecnologías, problemas y abordajes del estado del arte.



**Advisor**  
Mg. Rodolfo Bertone  
**Scientist Consultant**  
Dr. Hernán Merlino  
**Thesis approval date**  
May 17, 2022

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

## **Formulation of Processes for Spatial Knowledge Discovery Engineering**

**Keywords:** Knowledge discovery; Knowledge discovery process; Spatial Data; Georeferenced data; Spatial data mining.

### **Motivation**

Knowledge discovery processes provide a logical order and structure for the activities needed to extract non-trivial knowledge that is implicit in organizational data.

Spatial data are a type of contextual data that allows us to model events, objects or situations located in specific spatial locations. By using this kind of data it is possible to extract patterns that not only involve the data attributes but also the spatial relations between data entities, making the whole process more complex.

However, even if there are algorithms specialized in extracting these complex relations, multiple valuable conceptual aspects from spatial data mining are generally ignored in practice: traditional data mining algorithms are often used utilizing processes that were not designed to handle the particularities of this kind of information.

In consequence, it becomes necessary to develop and validate a set of knowledge discovery processes that systematize the activities that have to be used to extract knowledge from spatial data without ignoring their particularities. More specifically

- To identify kinds of patterns that can be extracted from spatial data.
- To identify spatial data properties that add value to the strategic decision-making process.
- To design knowledge discovery processes for the extraction of patterns from spatial data using techniques and solutions implemented in the technological tools that currently exist.

### **Thesis contributions**

Four spatial knowledge discovery processes have been developed:

- A process for the discovery of spatial groups that allows the user to find groups, regions and hotspots and describe them using the non-spatial attributes of the input data.
- A process for spatial anomaly discovery that enables the user to find local anomalies and group anomalies and describe their behavior in contrast to normal data.
- A process for spatial association discovery in local contexts considering multiple crisp and fuzzy spatial relations among the data entities.
- A process for co-location discovery in local contexts that can be extended to other kind of spatial relations.

All the aforementioned artifacts use vectorial data as input. Also, these processes consider the possibility of taking into account multiple crisp or fuzzy spatial relations, generally by using graph theory, and allow contemplating the spatial autocorrelation and spatial heterogeneity phenomena. Lastly, these processes are independent from specific algorithms and technologies.

### **Future Research Lines**

During the development of the aforementioned processes, the following concerns have come up and then have been relegated to future works:

- There is a wide variety of spatial patterns that have not been considered. Particularly, spatial classification, which is one of the tasks with the higher relevance in the problem domain.



- It is necessary a set of processes that use raster data as input, without having to convert the data format.
- The artifacts were designed to be general enough to adapt to a wide range of situations. However, some problems need to be addressed with more specific approaches. A systematic literature review is needed to map these problems with the algorithms and methods used to solve them.
- There are other spatial phenomena that have not been explored and might be interesting for decision making in some scenarios.
- Lastly, a periodic update of the processes is needed to include changes according to the new technologies and approaches from the state of the art.



**Director**  
Dr. Jorge Ierache

**Codirector**  
Dr. Waldo Hasperue

**Asesor Científico**  
Dr. Enrique Calot

**Fecha de aprobación**  
7 de junio de 2022

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

## **Generalización del Modelado de Cadencias de Tecleo con Contextos Finitos para su Utilización en Ataques de Presentación y Canal Lateral**

**Palabras clave:** Interacción hombre-máquina; Biometría comportamental; Seguridad de la información; Cadencias de tecleo; Aprendizaje automático

### **Motivación**

Las técnicas de análisis y modelado de la cadencia de tecleo han sido y continúan siendo ampliamente estudiadas por su interés intrínseco y para su aplicación como un segundo factor en la autenticación de usuarios. En particular, estos métodos tienen la ventaja de ser transparentes, en el sentido de que no demandan ninguna acción ulterior al usuario legítimo más que escribir naturalmente, y no le imponen demoras, confirmaciones, u otras tareas como los biométricos convencionales y demás esquemas de autenticación multifactor.

Sin embargo, las vulnerabilidades de los sistemas de autenticación basados en cadencias de tecleo ante ataques de presentación y falsificaciones sintéticas han recibido mucha menos atención que en el caso de los biométricos tradicionales, a pesar de que las debilidades a las que los expone no es inferior. Asimismo, el análisis y modelado de cadencias de tecleo es promisorio como vector de amplificación de ataques de canal lateral.

Es por lo tanto de interés para el progreso de la disciplina del análisis biométrico comportamental, enmarcada en el ámbito de la seguridad informática y de la interacción hombre-máquina, avanzar la investigación exploratoria de técnicas factibles de esclarecer ambos problemas: el ataque en contra y la defensa de los sistemas de autenticación basados en cadencias de tecleo. El modelado con contextos finitos promete brindar esclarecimiento teórico y soluciones prácticas dentro de este alcance al problema de la detección de vida, en el que los biométricos tradicionales se encuentran mucho más avanzados.

### **Los objetivos buscados en el presente estudio son:**

- Identificar las distribuciones subyacentes y los patrones de comportamiento que generan la cadencia de tecleo en texto libre de un usuario legítimo por medio de una comparación sistemática de histogramas empíricos de tiempos y diversos métodos de evaluación. No existe actualmente una comparación sistemática en texto libre.
- Proponer estrategias de síntesis de muestras artificiales para su uso en ataques de presentación que logren engañar a los actuales sistemas de autenticación basados en cadencias de tecleo con frecuencia suficiente como para constituir una amenaza.
- Evaluar un sistema de detección de vida que sirva como contramedida de defensa ante las anteriores y otras estrategias de síntesis de muestras artificiales.
- Explorar las técnicas derivadas de los anteriores para su uso en ataques de canal lateral pues el incremento en la capacidad para generar muestras sintéticas indistinguibles de las del usuario legítimo puede ser aprovechado en la identificación del texto ingresado.
- Crear una herramienta que permita integrar los resultados de las anteriores.

### **Aportes de la tesis**

Los aportes de esta tesis abarcan el conjunto de la producción científica, cuatro métodos novedosos, la herramienta integrada de evaluación, y los conjuntos de datos de evaluación y de resultados para los experimentos principales y secundarios.

- Producción científica
  - 2 artículos en revistas de primer cuartil
  - 6 artículos en revistas nacionales e internacionales
  - 6 artículos en workshops nacionales
- Métodos
  - Distancias CDF
  - Estrategia de síntesis ICDF
  - Detección de muestras sintéticas
  - Identificación de textos en base a los tiempos de escritura
- Herramienta
- Conjuntos de datos
  - 6 conjuntos de datos disponibles en IEEE DataPort y Mendeley Data.

### Líneas de I/D futuras

El objetivo de entender los procesos neuromusculares que hacen que los valores de tiempos al escribir se distribuyan como una log-logística puede abrir una nueva línea de investigación, que la interfaz cerebro-maquina que fue utilizada por miembros del Laboratorio de Sistemas de Información Avanzados [Calot 2019] puede ayudar a esclarecer. La estrategia NS/ICDF resultó peor que la más sencilla ICDF con entrenamiento intrausuario. Investigar qué otra estrategia alternativa puede mejorar la síntesis utilizando variaciones a nivel de palabras es una línea de investigación que se abre.

Los comportamientos individuales conforman un zoológico de idiosincrasias que quizás valga la pena discriminar en un estudio ulterior. La eficacia de las distancias basadas en histogramas empíricos para la tarea incentivan al análisis ulterior de qué otros atributos pueden extraerse de la cadencia de tecleo, que caractericen no al proceso motor general sino las especificidades de cada usuario.



**Advisor**  
Dr. Jorge Ierache

**Codirector**  
Dr. Waldo Hasperue

**Scientist Consultant**  
Dr. Enrique Calot

**Thesis approval date**  
June 7, 2022

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

## **Generalizing the Finite Context Modelling Method for Keystroke Dynamics to Presentation and Side-channel Attacks**

**Keywords:** Human-computer interaction; Behavioral biometrics; Information security; Keystroke dynamics; Machine learning

### **Motivation**

Keystroke dynamics analysis and modelling have been, and still are, widely studied both for their intrinsic interest and their applications as a second-factor authentication. In particular, these types of methods have the advantage of being transparent, in the sense that no further action is requested from the user other than his natural usage of the system. Unlike conventional biometrics and other multifactor authentication schemes, no delays, challenges, or tasks are imposed on the user.

However, vulnerabilities of keystroke dynamics authentication systems have received less attention than those of conventional biometrics. The risks include presentation attacks and synthetic forgeries, as well as the usage of keystroke timings leaked by a side-channel to boost a later brute-force attack.

Thus, it is important for the improvement of behavioral biometrics in general, and specifically of keystroke dynamics authentication systems, to explore the risks presented by the above attacks and how to defend against them. Finite context modelling can provide a practical solution to the challenges of liveness detection in this scope, in an attempt to match the recent achievements of conventional biometrics.

#### **The objectives of this study are:**

- Identify the underlying distributions of keystroke timings together with the characteristic behavioral patterns of keystroke dynamics, through a systematic comparison of empirical histograms using time-tested evaluation criteria. No systematic comparison of underlying distributions exists

in current literature for free-text keystroke dynamics.

- Propose methods for the synthesis of keystroke dynamics samples to be used in a presentation attack with the objective of fooling an existing keystroke dynamics authentication system.
- Evaluate a liveness detection method that can be used as a protection against the aforementioned threats.
- Explore how the techniques derived from the previous objectives can be used in side-channel attacks, as improves in the quality of synthetic samples that cannot be distinguished from those of the legitimate user can be leveraged to identify the typed text based on its keystroke timings only.
- Build a tool that integrates and simplifies the execution of the previous methods.

### **Thesis contributions**

The contributions of this thesis encompass the scientific production, four novel methods, the evaluation tool, and the datasets.

- Scientific production
  - 2 articles in Q1 journals
  - 6 articles in national and international journals
  - 6 articles in national workshops
- Methods
  - CDF-based distances
  - ICDF synthesis
  - Liveness detection for keystroke dynamics
  - Text identification based on keystroke timings

- Evaluation tool
- Datasets
- 6 datasets, available at IEEE DataPort and Mendeley Data repositories

### **Future Research Lines**

Attempts to understand the neuromuscular processes that determine keystroke timings can open a new line of research. The brain-machine interface pioneered by members of the Laboratorio de Sistemas de Investigación Avanzados [Calot 2019] can help to clarify why they follow a log-logistic distribution.

The synthesis method NS/ICDF underperformed in comparison with the much simpler ICDF when within-subject training was used. Finding out which method can improve the synthesis of samples by leveraging non-stationarity at word boundaries is an open research question.

Further studies should attempt to understand how user behavior can be clustered. The performance achieved by CDF-based distances incentivize the search for empirical attributes derived from keystroke dynamics, which characterize the specific behavior of each user instead of relying on features of the general typing process.



**Directores**  
Ing. Luis Marrone  
Dr. Fernando G. Tinetti

**Fecha de aprobación**  
27 de junio de 2022

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

## Cálculo Científico Distribuido sobre Clientes Móviles Indeterminados en redes MANETs

**Palabras clave:** MANET; Protocolo; Incertidumbre; Clustering; Cross-Layer

### Motivación

En este trabajo se propone el diseño de un nuevo protocolo que combina un enfoque para lidiar con la indeterminación de clientes al momento de comenzar un cálculo científico distribuido basado en un diseño cross layer con un mecanismo que agrupa los dispositivos presentes en una red móvil ad hoc o MANET (Mobile Ad hoc Networks) en clusters, sacando rédito de la heterogeneidad inherente a este tipo de redes a fin de determinar qué dispositivos serán los líderes de los agrupamientos conformados.

### Aportes de la tesis

Un nuevo protocolo llamado PROCCDIMI (Protocolo para el Cálculo Científico Distribuido sobre clientes Móviles Indeterminados) que organiza los nodos de la MANET en clusters y elige coordinadores de estos agrupamientos lógicos a los dispositivos con capacidades superiores, para que estos asuman el rol de asignar los clusters de pertenencia a las unidades de red que lo solicitan, tomar el cálculo científico entre los cliente que llevarán a cabo dicho cómputo en forma distribuida y finalmente recolectar todos los resultados parciales, compilarlos y devolver el resultado del cálculo científico. De los resultados obtenidos se arribaron a las siguientes conclusiones:

- PROCCDIMI es notoriamente superior en términos del Packet Delivery Ratio (PDR) y Throughput cuando se trata de redes relativamente pequeñas en cuanto a cantidad de nodos y cuando el número de nodos asciende exhibe un desempeño similar al resto de los protocolos de enrutamiento contra los cuales se lo comparó.
- PROCCDIMI es el mejor en cuanto al manejo de la utilización de la energía. Este hecho es producto de la estrategia de conformar grupos lógicos de nodos.

### Líneas de I/D futuras

A continuación, se resumen las principales líneas futuras de investigación:

- Probar el desempeño de PROCCDIMI en entornos más reales: Modelar la generación de tráfico en la red de forma más realista, para que no sea únicamente la información transmitida entre los nodos de la ruta establecida la que circule por la red, si no que pueda haber interferencias debidas a otras comunicaciones.
- Implementar un mecanismo de seguridad: Otro aspecto a evaluar en el protocolo PROCCDIMI es la incorporación de medidas que provean seguridad-privacidad.
- Implementar la automatización de elección del coordinador de cluster: En las simulaciones realizadas los escenarios definían estáticamente los tipos de dispositivos y esta definición era realizada ad hoc. Con el propósito de hacer más sólida y estricta la estrategia de elección de coordinador es preciso emplear algoritmos ampliamente probados como por ejemplo: Highest-Degree heuristic, Lowest-ID heuristic, Node-weight heuristic, Weighted Clustering algorithm o Distributed Weighted Clustering Algorithm.
- Probar PROCCDIMI con otro protocolo de enrutamiento: A la luz de los resultados obtenidos en la simulaciones y a la naturaleza intrínseca de los protocolos de enrutamiento que funcionan sobre las MANET, se puede decir que los protocolos reactivos introducen poca carga de paquetes de control, como resultado se optimiza el ancho de banda y se obtienen altas tasas de entrega de paquetes durante el proceso de descubrimiento de nuevos clientes; pero en cambio se registran tiempos de latencias muy elevados. Por tal razón resultaría de interés evaluar a PROCCDIMI con un protocolo de ruteo proactivo como sustrato para realizar toda su operatoria reemplazando a AODV (Ad hoc On-Demand Distance Vector), con el objetivo de dimensionar en qué medida mejoran los tiempos de latencias.



**Advisors**  
Ing. Luis Marrone  
Dr. Fernando G. Tinetti

**Thesis approval date**  
June 27, 2022

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

## Scientific Calculation Distributed on Undetermined Mobile Clients in MANETs networks

**Keywords:** MANET; Protocol; Uncertain; Clustering; Cross-Layer

### Motivation

In this work, the design of a new protocol is proposed that manages the uncertainty of clients at the time of starting a distributed scientific calculation. This new protocol is based on a cross layer design with a mechanism that groups the devices present in a MANET into clusters. Also, manages the heterogeneity of the devices that make up the wireless mobile network, so that the network units with higher computing power are the leaders of the groups.

### Thesis contributions

A new protocol called PROCCDIMI (Protocolo para el Cálculo Científico Distribuido sobre clientes Móviles Indeterminados) was proposed, which organizes the MANET nodes into clusters and chooses coordinators of these logical groupings for the devices with higher capacities, so that they assume the role of assigning the clusters belonging to the network units who request it, take the scientific calculation among the clients who will carry out this calculation in a distributed way and finally collect all the partial results, compile them and return the result of the scientific calculation. From the results obtained, the following conclusions have been reached:

- PROCCDIMI is notoriously superior in terms of the Packet Delivery Ratio (PDR) and Throughput when it comes to relatively small networks in terms of number of nodes and when the number of nodes increases it exhibits a similar performance to the rest of the routing protocols against which it is used. compared it.
- PROCCDIMI is the best in terms of managing the use of energy. This fact is the result of the strategy of forming logical groups of nodes.

### Future Research Lines

The main contributions to be made with this doctoral thesis are:

- Test the performance of PROCCDIMI in more real environments: Model the generation of traffic in the network in a more realistic way, so that it is not only the information transmitted between the nodes of the established route that circulates through the network, but that there may be interference due to other communications.

- Implement a security mechanism: Another aspect to be evaluated in the PROCCDIMI protocol is the incorporation of measures that provide security-privacy.

- Implement the automation for choosing the cluster coordinator: In the simulations carried out, the scenarios statically defined the types of devices and this definition was carried out ad hoc. In order to make the coordinator choice strategy more robust and strict, it is necessary to use widely proven algorithms such as: Highest-Degree heuristic, Lowest-ID heuristic, Node-weight heuristic, Weighted Clustering algorithm or Distributed Weighted Clustering Algorithm.

- Test PROCCDIMI with another routing protocol: In light of the results obtained in the simulations and the intrinsic nature of the routing protocols that work on MANETs, it can be said that reactive protocols introduce little load of control packets, as a result bandwidth is optimized and high packet delivery rates are achieved during the new customer discovery process; but on the other hand, very high latency times are recorded. For this reason, it would be of interest to evaluate PROCCDIMI with a proactive routing protocol as a substrate to carry out all its operations, replacing AODV (Ad hoc On-Demand Distance Vector), in order to determine the extent to which latency times improve.



**Directora**  
Dra. Marcela Genero Bocco

**Director**  
Mg. Rodolfo Alfredo Bertone

**Fecha de aprobación**  
7 de julio de 2022

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

## **DepProMod: Modelo de Proceso de Despliegue de Sistemas de Software**

**Palabras clave:** Despliegue de sistemas de software; Modelo de proceso; PyMES; Estudios de casos; Design science.

### **Motivación**

El despliegue de sistemas de software es un proceso crucial del ciclo de vida de desarrollo de software porque su resultado determina la aceptación exitosa o no del sistema de software por parte del cliente. En el despliegue se presentan inconvenientes como por ejemplo la falta de componentes (externos), descargas incompletas, la infraestructura heterogénea e incompatible, incumplimiento de calendarios, retrabajo por falta de pocas capacidades y competencias tecnológicas, capacitación y preparación inadecuadas de los usuarios finales entre otros puntualmente en Pequeñas y Medianas Empresas (PyMES).

En Argentina, la industria del software se compone mayoritariamente por PyMES, representando casi el 80% del sector, lo que constituye un eslabón fundamental, en el sector, para el país y refuerza la necesidad de llevar adelante iniciativas que contribuyan con el desarrollo y mejora de competitividad de dichas empresas.

En el contexto internacional se refleja la misma realidad respecto a que las PyMES ocupan una gran porción de la industria del software. Estas organizaciones descubrieron que es crucial mejorar sus procesos y métodos para ser competitivas en su negocio, aunque no cuentan con recursos y conocimiento suficiente para lograrlo. La única manera de lograr proyectos exitosos es mediante la definición, implementación y estabilización de los procesos de desarrollo.

Esta realidad motivó el objetivo de esta tesis doctoral, que es “La definición y validación de un modelo de proceso para realizar el despliegue de sistemas software desarrollados a medida de manera sistematizada y controlada en PyMES de Argentina denominado DepProMod (en inglés,

Deployment Process Model)”. Este ha sido creado, refinado y validado, siguiendo el marco metodológico Design Science (Ciencia del Diseño).

### **Aportes de la tesis**

#### **• DepProMod.**

Un modelo de proceso de despliegue de sistemas de software que le permita a las PyMES realizar el despliegue de sistemas de software desarrollados a medida de manera sistematizada y controlada. DepProMod propone fases, actividades y tareas, roles, entradas y salidas, plantillas y recursos y además cuenta con una estructura de tres niveles de aplicación para facilitar su implantación en PyMES. DepProMod se definió y se evaluó por diferentes profesionales de la industria del software para asegurar que fuera viable y útil. Este modelo también podría ser útil para los investigadores interesados en ampliar la investigación presentada en esta tesis doctoral. En el contexto de la academia, sería útil incorporarlo en los programas de las asignaturas para que los profesores puedan explicarlo como una solución alternativa para resolver los inconvenientes del proceso de despliegue de sistemas de software.

#### **• Conocimiento generado a través de los estudios empíricos.**

El conocimiento creado a través de varios estudios empíricos realizados durante el desarrollo de la tesis doctoral. Un mapeo sistemático de la literatura realizado junto con una encuesta sobre la práctica actual del despliegue de sistemas de software en PyMES de Argentina que permitieron elaborar el estado de arte. El primer estudio de caso llevado a cabo en una PyME desarrolladora de software de



Argentina que permitió examinar la viabilidad de aplicación de la versión preliminar de DepProMod para refinarlo y completarlo. Las entrevistas realizadas a dos expertos de la industria del software para conocer su opinión sobre los elementos del modelo y su utilidad. Los tres estudios de casos realizados en PyMES desarrolladoras de software de Argentina que permitieron evaluar la percepción de los usuarios de DepProMod medida a través de tres variables, la “Facilidad de Uso Percibida”, la “Utilidad Percibida” y la “Intención de Uso” y que además permitieron identificar las fortalezas y debilidades de DepProMod informadas por los usuarios del modelo.

### Líneas de I/D futuras

#### • Refinar DepProMod.

Mejorar la percepción de los usuarios con respecto a algunos elementos, como las plantillas. Para ello, será necesario la revisión, simplificación y en algún caso, eliminación de las plantillas como futura mejora de DepProMod.

#### • Continuar la investigación empírica.

- Evaluar el modelo en PyMES de otros sectores industriales que desarrollen software para su propio uso.

o Incorporar en los estudios de casos la recolección de datos cuantitativos como por ejemplo métricas que se desprendan del proceso de despliegue.

- Realizar más estudios de casos que sean llevados a cabo por otros investigadores.

- Extender la evaluación de DepProMod a un contexto internacional.

#### • Extender el alcance de DepProMod.

- Evaluar la viabilidad de poder aplicar el modelo en empresas emergentes (Startups) o empresas de mayor tamaño para recolectar feedback por parte de los usuarios y contrastarlo con los resultados obtenidos en PyMES.

- Evaluar la viabilidad de poder aplicar DepProMod en el despliegue de otro tipo de sistemas de software, por ejemplo, los sistemas de software configurables.



**Advisors**

Dra. Marcela Genero Bocco  
Mg. Rodolfo Alfredo Bertone

**Thesis approval date**  
July 7, 2022

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

**DepProMod: Deployment  
Process Model for  
software systems**

**Keywords:** Software system deployment; Process model; PyMES; Case studies; Design science

**Motivation**

The deployment of software systems is a crucial process of the software development life cycle because its outcome determines the successful acceptance or otherwise of the software system by the customer. In the deployment there are drawbacks such as the lack of (external) components, incomplete downloads, heterogeneous and incompatible infrastructure, non-compliance with schedules, rework due to lack of few technological capabilities and skills, inadequate training, and preparation of end users among others. punctually in Small and Medium Enterprises (SMEs).

In Argentina, the software industry is made up mostly of SMEs, representing almost 80% of the sector, which constitutes a fundamental link, in the sector, for the country and reinforces the need to carry out initiatives that contribute to the development and improvement competitiveness of these companies.

In the international context, the same reality is reflected that SMEs occupy a large portion of the software industry. These organizations discovered that it is crucial to improve their processes and methods to be competitive in their business, although they do not have sufficient resources and knowledge to do so. The only way to achieve successful projects is through the definition, implementation and stabilization of development processes.

This reality motivated the objective of this doctoral thesis, which is “The definition and validation of a process model to carry out the deployment of custom-made software systems in a systematic and controlled manner in SMEs in Argentina called DepProMod (Deployment Process Model)”. This has been created, refined, and validated, following the Design Science methodological framework.

**Thesis contributions**

• **DepProMod.**

A software system deployment process model that allows SMEs to deploy custom-developed software systems in a systematic and controlled manner. DepProMod proposes phases, activities and tasks, roles, inputs, and outputs, templates, and resources, and it also has a three-level application structure to facilitate its implementation in SMEs. DepProMod was defined and evaluated by different professionals in the software industry to ensure that it was viable and useful. This model could also be useful for researchers interested in expanding the research presented in this doctoral thesis. In the academic context, it would be useful to incorporate it into course syllabi so that teachers can explain it as an alternative solution to solving the drawbacks of the software system deployment process.

• **Knowledge created through empirical studies.**

The knowledge created through several empirical studies carried out during the development of the doctoral thesis. A systematic mapping study of the literature carried out together with a survey on the current practice of deploying software systems in SMEs in Argentina that allowed the elaboration of the state of the art. The first case study carried out in a software development SME from Argentina that allowed examining the feasibility of applying the preliminary version of DepProMod to refine and complete it. Interviews conducted with two experts from the software industry to find out their opinion on the elements of the model and its usefulness. The three case studies carried out in software development SMEs in Argentina that allowed evaluating the perception of DepProMod users measured through three variables, the “Perceived Ease of Use”, the “Perceived Utility” and the “Intention to Use” and that also allowed to identify

the strengths and weaknesses of DepProMod reported by the users of the model.

### Future Research Lines

- **Refine DepProMod.**

Improve the perception of users regarding some elements, such as templates. For this, it will be necessary to review, simplify and in some cases, eliminate the templates as a future improvement of DepProMod.

- **Continue the empirical research.**

- Evaluate the model in SMEs from other industrial sectors that develop software for their own use.
- Incorporate in the case studies the collection of quantitative data such as metrics that emerge from the deployment process.
- Carry out more case studies that are carried out by other researchers.
- Extend the evaluation of DepProMod to an international context.

- **Extend the scope of DepProMod.**

- Evaluate the feasibility of being able to apply the model in emerging companies (Startups) or larger companies to collect feedback from users and contrast it with the results obtained in SMEs.
- Evaluate the feasibility of being able to apply DepProMod in the deployment of other types of software systems, for example, configurable software systems.

## DOCTORADO EN CIENCIAS INFORMÁTICAS



**Dra. Gladys Noemí Kaplan**

e-mail  
gkaplan@unlam.edu.ar

español

**Director**  
Ing. Jorge Horacio Doorn

**Fecha de aprobación**  
11 de julio de 2022

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

## Proceso de Requisitos Validado Empíricamente

**Palabras clave:** Ingeniería de Requisitos; Proceso de Requisitos basado en Escenarios; Léxico Extendido del Lenguaje (LEL); Escenarios; Puntos de Vista; LEL de Requisitos; Información Extemporánea.

### Motivación

La participación de la tesista en la construcción del Proceso de Requisitos basado en Escenarios y la experiencia adquirida durante su uso en casos reales y la transferencia realizada en el aula ha permitido conocer sus virtudes y detectar algunas debilidades. Es justamente en estas debilidades en lo que se ha centrado el estudio empírico de la presente tesis, identificando las causas de las deficiencias, para luego proponer soluciones a todas ellas. El objetivo de este trabajo es, en primera instancia, asegurar la calidad de cada modelo utilizado (LEL y escenarios) para luego, garantizar la calidad de todo el proceso de requisitos. Cabe mencionar que los requisitos del software se obtienen de los escenarios futuros, los cuales se construyen tomando información de los modelos previos (LEL y escenarios actuales), por lo tanto, la calidad de cada modelo construido repercute directamente en la calidad de los requisitos de software obtenidos.

### Aportes de la tesis

Se puede sintetizar en tres puntos: 1) el Proceso de Requisitos basado en Escenarios utilizado también como marco teórico, en el cual la tesista ha tenido diferentes tipos de intervención durante más de dos décadas. 2) el estudio empírico realizado al Proceso de Requisitos en general y en particular a los modelos LEL y Escenarios y 3) un conjunto de agregados y mejoras que aumentan significativamente la calidad de los modelos utilizados y de todo el proceso en general, como ser dos nuevas heurísticas, una de construcción del LEL y otra para la Derivación, la inclusión de los puntos de vista del contexto en el LEL y en los escenarios, la inclusión de las jerarquías conceptuales del contexto, un mejor tratamiento de los estados, la incorporación del LEL de Requisitos y de la Información

extemporánea y algunas propuestas para mejorar la vinculación entre el LEL y los escenarios. Todas estas incorporaciones han permitido obtener requisitos de software más completos, reales y consistentes.

### Líneas de I/D futuras

Se espera probar las heurísticas nuevas en nuevos casos reales y compararlas con las heurísticas existentes. Con estos datos, de ser necesario, refinar aún más las heurísticas.



english

**Advisor**  
Ing. Jorge Horacio Doorn

**Thesis approval date**  
July 11, 2022

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

## **Proceso de Requisitos Validado Empíricamente**

**Keywords:** Requirements Engineering; Scenario-based Requirements Process; Extended Language Lexicon (LEL); Scenarios; Points of view; LEL of Requirements; Extemporaneous Information

### **Motivation**

The participation of the thesis student in the construction of the Requirements Process based on Scenarios and the experience acquired during its use in real cases and the transfer carried out in the classroom has allowed knowing its virtues and detecting some weaknesses. It is precisely on these weaknesses that the empirical study of this thesis has focused, identifying the causes of the deficiencies, and then proposing solutions to all of them.

The objective of this work is, in the first instance, to ensure the quality of each model used (LEL and scenarios) and then to guarantee the quality of the entire requirements process. It is worth mentioning that the software requirements are obtained from future scenarios, which are built taking information from previous models (LEL and current scenarios), therefore, the quality of each built model has a direct impact on the quality of the software requirements obtained software.

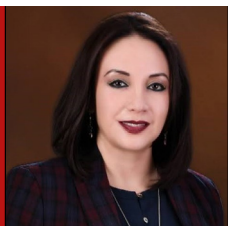
### **Thesis contributions**

It can be summarized in three points: 1) the Scenario-based Requirements Process also used as a theoretical framework, in which the thesis student has had different types of intervention for more than two decades. 2) the empirical study carried out on the Requirements Process in general and in particular on the LEL models and Scenarios and 3) a set of additions and improvements that significantly increase the quality of the models used and of the entire process in general, such as two new heuristics, one for the construction of the LEL and another for Derivation, the inclusion of the context viewpoints in the LEL and in the scenarios, the inclusion of the conceptual hierarchies of the context, a better treatment of the states, the incorporation of the

LEL of Requirements and extemporaneous Information and some proposals to improve the link between the LEL and the scenarios. All these additions have made it possible to obtain more complete, real and consistent software requirements.

### **Future Research Lines**

It is expected to test the new heuristics in new real cases and compare them with the existing heuristics. With these data, if necessary, further refine the heuristics.



**Directora**

Dra. Marisa Raquel De Giusti

**Codirector**

Dr. Fernando Molina Granja

**Asesor Científico**

Dr. Gonzalo Luján Villarreal

**Fecha de aprobación**

20 de septiembre de 2022

**SEDICI**

[Tesis Completa](#)

## **Modelo de madurez para preservación digital a largo plazo aplicando principios de planificación estratégica**

**Palabras clave:** Preservación digital; modelo de madurez; planificación estratégica; conservación de documentos.

### **Motivación**

La conservación de la información es una necesidad del ser humano, que se ha presentado durante el transcurso de la historia. Las organizaciones actualmente dependen en gran medida de la preservación digital, la gestión documental y la memoria corporativa. El conocimiento que esta genera, les permite a las organizaciones ser más competitivas, accediendo a una base de conocimiento histórico y científico. En la preservación digital se presentan retos técnicos como la migración de información, pérdida de datos, documentos con problemas de lectura, la accesibilidad de la información en el futuro, la unicidad de documentos y autenticidad. Por esta razón se han creado múltiples estándares, modelos de referencia, auditoría y madurez de preservación digital, que proporcionan importantes bases y guías para apoyar la implementación de proyectos en este ámbito, donde un aspecto que se debe considerar es el jurisdiccional, esto significa que en espacios legales, en muchas ocasiones no es suficiente cumplir a cabalidad un modelo o estándar, sino que también es posible que existan reglas impuestas por el país o lugar en el cual la información digital debe ser preservada. Además, es sustancial analizar los factores políticos, económicos, sociales, tecnológicos y ecológicos, que se encuentran alrededor de una organización.

El término madurez, indica la mejora de un sistema, proceso o procedimiento, que ha alcanzado su etapa de gestión más alta, por lo que se espera que, en términos de preservación digital, el departamento o unidad encargado

de realizar este trabajo, también alcance el mejor nivel de conservación de la información a largo plazo. El planteamiento de una madurez a largo plazo, está íntimamente relacionada a la planificación estratégica, debido a que, para alcanzarla, la organización debe establecer lineamientos claros, plantear su misión, observar su contexto, tanto interno como externo, plantear su visión, metas y objetivos a cumplirse en los años posteriores, hasta llegar a un alto nivel de madurez de preservación digital.

**Objetivo General:**

Desarrollar un modelo de madurez para preservación digital a largo plazo, aplicando principios de planificación estratégica.

**Objetivos Específicos:**

- Definir las características fundamentales de la preservación digital a largo plazo, a partir del estudio de sus estándares, modelos de auditoría y modelos de madurez.
- Adaptar las bases de la planificación estratégica a la implementación de proyectos de preservación digital a largo plazo.
- Diseñar un modelo de madurez para preservación digital basado en características conceptuales y prácticas de preservación digital, además de las bases de la planificación estratégica.
- Analizar la validez del modelo de madurez para preservación digital propuesto, mediante la evaluación de expertos con respecto a la preservación digital, planificación estratégica y a través de la aplicación de un caso de estudio.

## Aportes de la tesis

El aporte principal de esta investigación es el diseño de un modelo de madurez para preservación digital a largo plazo, aplicando principios de planificación estratégica. Esta investigación no pretende únicamente determinar aspectos referenciales, acerca de las fases o niveles en el proceso de implementación de un proyecto de preservación digital y las actividades relacionadas en cada una de ellas, sino, además, busca establecer un modelo de madurez con un proceso sistemático de desarrollo e implementación de actividades.

Además, el modelo incluye mecanismos de evaluación cuantitativa porcentual, seguimiento, medición de resultados, análisis por perspectivas, por objetivos estándares y mejora continua de los procesos, lo cual permite formular, implantar, dirigir y evaluar decisiones interfuncionales, que permitan a la organización alcanzar la preservación digital a largo plazo. Adicionalmente, existen aportes significativos como múltiples gráficos y tablas comparativas de estándares, modelos de auditoría, madurez de preservación digital y modelos de planificación estratégica, con sus características principales, fases y criterios a considerar en cada una de ellas. Y los instrumentos de aplicación del modelo en donde se incluye una matriz y una guía para la implementación.

## Líneas de I/D futuras

La investigación propuesta en este trabajo, abre varios ámbitos de estudio futuros, el primer ámbito de trabajo será la aplicación de este modelo de madurez en otros organismos que realicen preservación digital, con la finalidad de corroborar los resultados obtenidos en esta tesis doctoral. Además, el segundo ámbito, será la creación de investigaciones, para generar herramientas informáticas, que automaticen aún más la aplicación del proceso planteado. El tercer ámbito de investigación, será comprobar que el modelo propuesto, puede adaptarse, con ciertos cambios, para evaluar la madurez de sistemas de información en empresas, organismos y departamentos, permitiendo la mejora continua de los mismos a través de los principios de planificación estratégica.



**Advisor**  
Dra. Marisa Raquel De Giusti

**Codirector**  
Dr. Fernando Molina Granja

**Scientist Consultant**  
Dr. Gonzalo Luján Villarreal

**Thesis approval date**  
September 20, 2022

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

## **Long-term digital preservation maturity model applying strategic planning principles**

**Keywords:** Digital preservation; Maturity model; Strategic planning; Document preservation

### **Motivation**

Keystroke dynamics analysis and modelling have been, Information preservation is a human need that has been present throughout history. Organizations today rely heavily on digital preservation, document management and corporate memory. The knowledge it generates allows organizations to be more competitive by accessing a historical and scientific knowledge base. Digital preservation presents technical challenges such as information migration, data loss, documents with reading problems, accessibility of information in the future, uniqueness of documents and authenticity. For this reason, multiple standards, reference models, auditing and digital preservation maturity have been created, which provide important bases and guidelines to support the implementation of projects in this area, where one aspect that must be considered is the jurisdictional, this means that in legal spaces, in many cases it is not enough to fully comply with a model or standard, but it is also possible that there are rules imposed by the country or place in which the digital information must be preserved. In addition, it is important to analyze the political, economic, social, technological and ecological factors that surround an organization.

The term maturity indicates the improvement of a system, process or procedure, which has reached its highest management stage, so it is expected that, in terms of digital preservation, the department or unit in charge of performing this work, also reaches the best level of information preservation in the long term. The approach to long-term maturity is closely related to strategic planning, because, to

achieve it, the organization must establish clear guidelines, set out its mission, observe its context, both internal and external, set out its vision, goals and objectives to be met in subsequent years, until reaching a high level of digital preservation maturity.

#### **General Objective:**

To develop a maturity model for long-term digital preservation, applying strategic planning principles.

#### **Specific Objectives:**

- Define the fundamental characteristics of long-term digital preservation, from the study of its standards, audit models and maturity models.
- Adapt strategic planning principles to the implementation of long-term digital preservation projects.
- Design a maturity model for digital preservation based on conceptual and practical characteristics of digital preservation, in addition to the bases of strategic planning.
- Analyze the validity of the proposed digital preservation maturity model through expert assessment of digital preservation, strategic planning, and through the application of a case study.

### **Thesis contributions**

The main contribution of this research is the design of a maturity model for long-term digital preservation, applying strategic planning principles. This research is not only intended to determine referential aspects, about the phases or levels in the implementation process of a digital preservation project

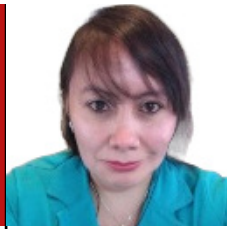


and the activities related to each of them, but also seeks to establish a maturity model with a systematic process of development and implementation of activities.

In addition, the model includes mechanisms for quantitative percentage evaluation, monitoring, measurement of results, analysis by perspectives, by standard objectives and continuous improvement of processes, which allows formulating, implementing, directing and evaluating cross-functional decisions that enable the organization to achieve long-term digital preservation. In addition, there are significant contributions such as multiple graphs and comparative tables of standards, audit models, digital preservation maturity and strategic planning models, with their main characteristics, phases and criteria to be considered in each of them. And the tools for applying the model, including a matrix and a guide for implementation.

### **Future Research Lines**

The research proposed in this work opens several areas of future study, the first area of work will be the application of this maturity model in other organizations that perform digital preservation, in order to corroborate the results obtained in this doctoral thesis. In addition, the second area will be the creation of research to generate software tools to further automate the application of the proposed process. The third area of research will be to verify that the proposed model can be adapted, with certain changes, to evaluate the maturity of information systems in companies, organizations and departments, allowing their continuous improvement through the principles of strategic planning.



**Directores**  
Dr. Matías Urbietta  
Dr. Gustavo Rossi

**Fecha de aprobación**  
26 de octubre de 2022

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

## **Interfaces Proactivas Móviles desde Legacies Web: Un Enfoque de Diseño Conductual**

**Palabras clave:** Aplicación legacy; Cadenas de Markov; Enfoque empático; Interfaz proactiva; Mapeo sistemático; Migración; Dispositivos móviles; Portabilización.

### **Motivación**

En esta tesis se propone un enfoque empático multidisciplinario que permite la correcta portabilización de legacies web a móviles a través de algoritmos predictivos como las cadenas de Markov que ofrecen una mejorada experiencia de usuario a través de la obtención del modelo de comportamiento de navegación avalada por métricas de usabilidad, de manera adicional se logra la mejora de productividad laboral y el cumplimiento de las heurísticas de Nielsen. Los principales objetivos fueron:

- Investigar los métodos utilizados para la portabilización de aplicaciones legacies a móviles mediante una revisión sistemática de la literatura de la última década.
- Establecer una metodología para el análisis de aplicaciones Web que permita una correcta portabilidad a dispositivos móviles a partir de la identificación de necesidades del usuario y del negocio integrando los componentes y servicios más adecuados para la versión móvil.
- Desarrollar una herramienta que permita implementar de forma práctica el enfoque propuesto para evidenciar la factibilidad técnica.
- Evaluar la metodología con su aplicación en casos de estudio utilizando métricas de usabilidad, productividad y accesibilidad.

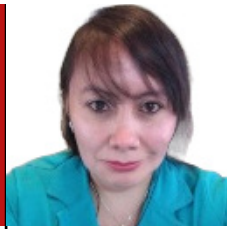
### **Aportes de la tesis**

La contribución incluye un trabajo transdisciplinario que demuestra que este modelo hace posible un diseño de interfaz inteligente y proactivo, considerando además variables sociodemográficas directamente relacionadas que contribuyen a la economía del contexto. Este estudio apoya la idea de que incluir cadenas de Markov para predecir el comportamiento del usuario podría ser una de las teorías conductuales que se pueden aplicar para mejorar otros enfoques, como los sistemas basados en modelos, en términos de asignación automática de pesos y prioridades. Por otro lado la tecnología al fin de cuentas debe apoyar la producción de valor en el negocio, por lo que los enfoques para el desarrollo y diseño de sistemas deben considerar metodologías interdisciplinarias, transversales que incluyan profesionales o teorías establecidas por otras áreas como la sociología, economía, matemática que permitan hu-

manizar el software con la capacidad de predecir las conductas para que el producto llegue a ser lo que el cliente quiere, inclusive alcanzar una investigación aplicada al ciudadano. A través de este tipo de enfoques, las empresas pueden llegar a ser más independientes de las redes sociales enriqueciendo la base de clientes, generando su propio big data, CRMs, business intelligence y finalmente inbound marketing que permita ser una organización orientada por los datos generando la hiperpersonalización. Por tanto, los perfiles de usuario juegan un papel importante ya que ayudan a definir diferentes niveles de personalización para aumentar el valor de vida útil del cliente (CLV) y también el retorno sobre inversión (ROI) en la transformación digital.

### **Líneas de I/D futuras**

- Como trabajo futuro, se planea la mejora continua del enfoque e integrar en primera instancia todas las personalizaciones en un solo sitio web para que el usuario con un único inicio de sesión, pueda acceder a todas sus personalizaciones y comportamientos de navegación para que no deba instalar el script de un dispositivo a otro.
- Se propone la mejora continua de la herramienta que respalda este enfoque con el fin de incrementar sus beneficios, a través de un experimento continuo y controlado para evaluar su proactividad de acuerdo con el uso del sistema a convertirse en un software evolutivo.
- Incluir test complementarios y métricas adicionales relacionadas con la experiencia del usuario, como emociones, actitudes, pensamientos, comportamientos y percepciones de los usuarios.
- Se propone, además, identificar otras teorías o algoritmos para determinar el comportamiento de usuarios en sitios Web para resolver el problema de la demanda en los diseños empáticos de software.
- Finalmente es importante proponer una metodología para la visualización de grandes volúmenes de datos de legacies en móviles con una adecuada infografía también acorde a un enfoque empático.



**Advisors**  
Dr. Matías Urbieto  
Dr. Gustavo Rossi

**Thesis approval date**  
October 26, 2022

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

## **Mobile Proactive Interfaces from Web Legacies: A Behavioral Design Approach**

**Keywords:** Legacy application; Markov chains; Empathic approach; Proactive interface; Systematic mapping; Migration, Mobile devices; Portabilit.

### **Motivation**

In this thesis, a multidisciplinary empathic approach is proposed that allows the correct portability of web legacies to mobiles through predictive algorithms such as Markov chains that offer an improved user experience by obtaining the navigation behavior model supported by metrics. of usability, additionally the improvement of labor productivity and the fulfillment of the Nielsen heuristics are achieved. The main objectives were:

- To investigate the methods used to port legacy applications to mobiles through a systematic review of the literature of the last decade.
- Establish a methodology for the analysis of Web applications that allows correct portability to mobile devices based on the identification of user and business needs, integrating the most appropriate components and services for the mobile version.
- Develop a tool that allows the proposed approach to be implemented in a practical way to demonstrate technical feasibility.
- Evaluate the methodology with its application in case studies using usability, productivity and accessibility metrics.

### **Thesis contributions**

The contribution includes a transdisciplinary work that demonstrates that this model makes possible an intelligent and proactive interface design, also considering directly related sociodemographic variables that contribute to the economy of the context. This study supports the idea that including Markov chains to predict user behavior could be one of the behavioral theories that can be applied to improve other approaches, such as model-based systems, in terms of automatic assignment of weights and priorities. On the other hand, technology must ultimately support the production of value in the business, so the approaches to the development and design of systems must consider interdisciplinary methodologies, transversal that include

professionals or theories established by other areas such as sociology, economics, mathematics that allow humanizing the software with the ability to predict behaviors so that the product becomes what the client wants, including reaching an applied research to the citizen. Through this type of approach, companies can become more independent of social networks, enriching the customer base, generating their own big data, CRMs, business intelligence and finally inbound marketing that allows them to be a data-driven organization generating hyperpersonalization.

### **Future Research Lines**

- As future work, it is planned to continuously improve the approach and integrate in the first instance all the customizations in a single website so that the user with a single login can access all their customizations and browsing behaviors so that they do not have to install the script from one device to another.
- The continuous improvement of the tool that supports this approach is proposed in order to increase its benefits, through a continuous and controlled experiment to evaluate its proactivity according to the use of the system to become an evolutionary software.
- Include complementary tests and additional metrics related to the user experience, such as emotions, attitudes, thoughts, behaviors and user perceptions.
- It is also proposed to identify other theories or algorithms to determine the behavior of users on websites to solve the demand problem in empathic software designs.
- Finally, it is important to propose a methodology for the visualization of large volumes of legacy data on mobile phones with an adequate infographic, also according to an empathic approach.

## DOCTORADO EN CIENCIAS INFORMÁTICAS



**Dr. Jorge Geovanni Raura Ruiz**

e-mail

georaura@gmail.com

español

**Director**  
Dr. Oscar Dieste

**Codirector**  
Dra. Claudia Pons

**Fecha de aprobación**  
29 de noviembre de 2022

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

## Impacto de las características personales de los programadores en la efectividad de Test- Driven-Development (TDD)

**Palabras clave:** Test driven development; Ingeniería de software empírico; Factores humanos; Desarrollo de software

### Motivación

**Contexto:** El desarrollo dirigido por pruebas (Test Driven Development - TDD), es una estrategia de programación propuesta por Kent Beck (Beck,2002) como alternativa al desarrollo de software tradicional, ha sido una técnica ampliamente estudiada en la ingeniería de software experimental, especialmente con la realización de estudios que intentan demostrar su efectividad en términos de calidad del código y productividad de los programadores. No obstante, los resultados obtenidos en varios estudios muestran resultados que hasta el momento no son concluyentes.

El desarrollo de software es un proceso centrado en la persona. los aspectos humanos desempeñan, en consecuencia, un papel importante y han sido investigados en prácticamente todas las actividades de la Ingeniería de Software, teniendo un impacto igual o incluso superior a los factores técnicos. Un factor que puede explicar los resultados aparentemente contradictorios sobre la efectividad de TDD es la falta de consideración de las características personales de los programadores.

**Objetivo:** En este trabajo se propone la realización de una familia de experimentos (Basili,1999), para determinar la influencia de factores personales en la Calidad externa del código y en la Productividad de los programadores al aplicar TDD en comparación con el desarrollo iterativo con pruebas al final (ITLD).

### Aportes de la tesis

Creemos que uno de los principales aportes de nuestro estudio, que lo consideramos de carácter exploratorio, es haber comprobado cómo algunos aspectos personales como la motivación de los participantes de los experimentos (sean profesionales o estudiantes), incide en su interés al realizar las tareas experimentales y por tanto influye en su productividad. También observamos que la edad y el grado de completitud o cantidad de código entregado por los participantes al realizar las tareas experimentales fue un factor que influyó significativamente en su productividad, independientemente

de la técnica utilizada (TDD o ITLD). Además, la experiencia en el lenguaje de programación y el conocimiento del entorno de desarrollo son factores que también incidieron significativamente en los resultados obtenidos

### Líneas de I/D futuras

A nuestro criterio, los aspectos humanos que han sido considerados aún se encuentran en una fase primaria de estudio, por lo que se plantean las siguientes futuras líneas de investigación:

Replicar los experimentos considerando los factores humanos que hemos identificado en la investigación. En este caso, la selección de sujetos y asignación a grupos podría realizarse de forma segmentada (por valores/niveles de los factores humanos). Como consecuencia de este cambio, la investigación pasaría de exploratoria a confirmatoria.

Otro aspecto que debería estudiarse con mayor detalle es si la motivación de los desarrolladores tiene una correlación con la edad y, por tanto, con su interés en realizar las tareas experimentales adecuadamente. Como pudimos apreciar, los sujetos de mayor edad fueron los que menos se esforzaron por cumplir las tareas adecuadamente y esto influyó en los resultados obtenidos. Esta investigación no tendría una relación directa con TDD, pero sí con la investigación experimental en Ingeniería de Software.



english

**Advisor**  
Dr. Oscar Dieste  
**Codirector**  
Dra. Claudia Pons  
**Fecha de aprobación**  
November 29, 2022  
**SEDICI**  
[Full Text](#)

## Impact of programmers' human factors on the effectiveness of TestDriven-Development (TDD)

**Keywords:** Test driven development; Empirical software engineering; Human factors; Software development;

### Motivation

**Context:** Test Driven Development (TDD) is considered as a programming strategy, it is proposed by Kent Beck (Beck, 2002) as an alternative to the traditional software development process. It is a widely studied technique in experimental software engineering, especially in studies that aim to demonstrate its effectiveness in terms of code quality and programmer's productivity. However, the results from those studies show results that are not conclusive so far.

Software development is a human-centered process. Therefore, human aspects play an important role in the process of producing software and are an important topic of research in Software Engineering activities, those aspects have an impact that is equal to or even greater than technical factors. One factor that may explain the apparently contradictory results on the effectiveness of TDD is the lack of consideration of the personal characteristics of the programmers.

**Objective:** This work proposes the realization of a family of experiments to determine the influence of personal factors in the external quality of the code and in the productivity of programmers when applying TDD and the iterative development with tests at the end (ITLD).

### Thesis contributions

We believe that the main contribution of our study, which we consider to be of an exploratory nature, is the verification of how some personal aspects, such as the motivation of the participants in the experiments (whether they are professionals or students), affect their interest in carrying out the experimental tasks and therefore they influence their productivity. We also observed that, the age, the degree of completeness or the amount of code implemented by the participants are factors that significantly influence their productivity, regardless of

the technique (TDD or ITLD). In addition, experience in the programming language and knowledge of the development environment are factors that also had a significant impact on the results.

### Future Research Lines

From our personal standpoint, we consider that the human aspects that we considered are still in a primary phase of study, therefore we suggest the following future lines of research:

Replicate the experiments considering the human factors that we identified in the investigation. In this case, the selection of subjects and the assignment to groups could be performed by values, or by levels of human factors. Consequently, the research would go from exploratory to confirmatory.

Another aspect that should be studied in greater detail is the developers' motivation correlation with age and, therefore, with their interest in performing the experimental tasks properly. According to our observations, the older subjects were the ones who made the least effort to perform the tasks properly, and that had a direct impact in the results. In the end, that research would not have a direct relationship with TDD, but with experimental research in Software Engineering.

## DOCTORADO EN CIENCIAS INFORMÁTICAS



**Dra. Mónica del Rocío  
Romero Pazmiño**

e-mail

monica.romerop@info.unlp.edu.ar

**Director**  
Lic. Francisco Javier Díaz

**Asesora Científica**  
Mg. Ivana Harari

**Fecha de aprobación**  
6 de diciembre de 2022

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

## **Hope Project: Realidad Aumentada para Procesos de Enseñanza Aprendizaje en Niños con TEA**

**Palabras clave:** Autismo (ASD); Diseño Centrado en Usuario (DCU); Proyecto Esperanza; Prototipo de Software; Procesos Enseñanza y aprendizaje; Realidad Aumentada (RA); Trastorno De Espectro Autista (TEA), Tecnologías de la Información y Comunicación (TIC).

### **Motivación**

El autismo se define como una característica en el cual se producen variaciones en diferentes aspectos, como una disminución en la comunicación, imaginación y adaptación social. Las personas con Trastorno de Espectro Autista (TEA) se desenvuelven y experimentan de modos particulares el proceso educativo, muchos de los tratamientos que se utilizaron desde hace décadas como propuestas educativas van resultando obsoletos con el paso del tiempo, existen recursos tecnológicos basados en Realidad aumentada (RA) diseñados específicamente para posibilitar la accesibilidad de las personas con discapacidad. Este trabajo se enfoca en estos recursos tecnológicos dadas sus pertinentes características que los convierten en altamente idóneos para ser utilizadas en entornos de enseñanza y aprendizaje. Por tanto, se intenta profundizar en este campo y ser un proyecto que sume y aporte al escenario investigativo, el objetivo de esta tesis doctoral fue el análisis, diseño, desarrollo, implementación y evaluación del impacto de un prototipo de software que fue concebido desde un diseño inclusivo centrado en el usuario basado en RA y orientado al fortalecimiento de las capacidades de aprendizaje social y desarrollo cognitivo de la población infantil TEA. Este proceso científico abarcó varios años de investigación y desarrollo siendo dividido en etapas, con el fin de lograr el objetivo final propuesto. Para la consecución de este objetivo se planificó trabajar en tres aristas principales como: la exploración del papel de las Tecnologías de Información y Comunicación (TIC), el análisis del potencial de accesibilidad y usabilidad de un conjunto de aplicaciones educativas desarrolladas para población TEA, además de analizar, diseñar, implementar y validar el impacto y la efectividad del prototipo de software, desde la

perspectiva de diseño centrado en usuario (DCU). Dentro de la investigación académica puede verse y reconocerse poca literatura y poca incidencia a ella relacionada, respecto al tema que el presente trabajo aborda. Si bien es cierto, la RA apoya el aprendizaje, miramos que, en la población infantil TEA no se implementa soluciones con la frecuencia y con la intensidad deseable y requerida, lo que justifica que se explore aún más sobre las ventajas de tales escenarios. La investigación no es sólo, pues, una investigación teórica y exploratoria. Este trabajo aplica los conocimientos alrededor del uso de la RA, y construye una aplicación que se suma al abanico de posibilidades de los agentes educativos en estos escenarios. Así, este trabajo intenta profundizar sobre un campo que, si bien todavía está poco explorado, sí presenta índices prometedores de crecimiento, de manera que este trabajo puede verse como un aporte importante en esta suma constante por la búsqueda de mejores condiciones de vida para la población infantil TEA. Este trabajo arroja información relevante que permite, primero, dar a conocer las posibilidades tecnológicas dentro de los entornos educativos y, segundo, incidir positivamente en la práctica y en el análisis científico acerca de las posibilidades, ventajas y desventajas de la implementación de la RA en procesos educativos con población infantil TEA.

### **Aportes de la tesis**

Esta investigación permitió analizar, investigar, desarrollar, implementar y experimentar mediante la creación de un prototipo que usa RA como herramienta pedagógica que complementa las actividades educativas de la población infantil TEA, que, por medio de un diseño inclusivo y basán-



dose en un DCU incluyó criterios de calidad, usabilidad y funcionalidad diferenciales frente al actual estado de aplicaciones orientadas a niñez TEA y basadas en RA.

Este es un proyecto de innovación en la medida que integra varias ramas del saber cómo: interacción hombre computador, accesibilidad, diseño centrado en el usuario, destinado para ofrecer inclusión personas con discapacidad. Realizó una interacción sistémica entre personal docente, personal médico, personal TIC, población infantil y familia TEA, además de incluir a un equipo multidisciplinario.

Se definió el Tratamiento Educativo de Niños con ASD mediado a través de RA denominado TEARA, que se diseñó para fortalecer procesos de enseñanza y aprendizaje promoviendo el uso de la vista, el oído, la percepción del cuerpo y el movimiento para niños con TEA leve, moderado, o severo puede ser utilizado por niñez desde los 4 años hasta los 16 años.

TEARA utiliza un sistema de entrenamiento denominado Software Hope, que enseña expresiones artísticas a través de la danza. Este sistema permite la movilidad de la población infantil TEA además de ofrecer opciones y configurar el entorno lúdico agregando o eliminando opciones para su uso, busca de forma progresiva incluir nuevos aprendizajes, mostrando a medida que pasa el tiempo una mayor complejidad, puede ser utilizado en centros académicos, centros de terapia, o en el domicilio.

El diseño del software denominado Hope no fue conductual permitió que la población infantil TEA aprenda a su ritmo; es un trabajo original, ya que a nivel mundial no se ha creado software que usando la RA enseñen danza a población infantil con TEA, generamos nuevos conocimientos ya que en la actualidad no existe fuente bibliográfica de diseño inclusivo centrado en el usuario además de aportar con guías para ejecutar pruebas de usabilidad para población infantil con TEA utilizando RA.

Evaluamos las condiciones existentes para la implementación de la RA en el proceso de enseñanza y aprendizaje de población infantil, diseñamos y validamos del prototipo tecnológico, destinado a contextos educativos

además de definir una Metodología de la intervención del software diseñado con RA. Profundizamos en una etapa de reflexión y discusión sobre la experiencia vivida basados en la experimentación en cada una de las fases. Finalmente, la utilización de las TIC y de RA se realizó de una manera activa, propositiva pero no intrusiva, permitió derribar ciertas creencias donde se indica que la tecnología de RA son adictivas y por tanto no son aconsejables para la población infantil TEA, demostramos con bases a los resultados que la utilización controlada mediante un plan de intervención previamente definido y con el apoyo de un equipo multidisciplinario puede aportar de manera substancial al aprendizaje de procesos como la imitación, percepción, motricidad fina y gruesa, además de coordinación viso -motriz.

## Líneas de I/D futuras

La investigación doctoral conllevó años de investigación, sobre la interacción población infantil TEA y tecnologías de RA. Por otra parte, entre las líneas de investigación que se pueden abrir a partir de este estudio se pueden mencionar las siguientes:

- Difundir TEARA como una forma disruptiva de innovar las estrategias de intervención en procesos de enseñanza aprendizaje que partan de experiencias integradoras para el acompañamiento en procesos de enseñanza y aprendizaje de los niños con autismo.
- Considerar la creación de programas de capacitación donde se difunda TEARA en la población (personal docente, médico y rehabilitador) y los beneficios de utilizar el software de entrenamiento Hope para la educación de población infantil TEA, explicando los elementos de hardware y software que se requieren en la intervención.
- Implementar espacios lúdicos basadas en la TEARA, en unidades educativas o centros de atención con el fin de que puedan intervenir otros tipos de discapacidades para fortalecer procesos de enseñanza y aprendizaje a través del tratamiento educativo propuesto.
- Ampliar el estudio para proveer al Sistema de entrenamiento Hope de nuevos procesos de enseñanza aprendizaje adicionales a los establecidos (imitación, percepción, motricidad fina y gruesa, coordinación viso -motriz) se puede ampliar a procesos que incluyan la flexibilidad del pensamiento memoria e imaginación.
- TEARA puede ser utilizado para verificar procesos de enseñanza aprendizaje en adolescentes y adultos mayores TEA. Ampliar la propuesta de intervención en niños con autismo leve, es posible que este plan y sus resultados sean mejor recibidos y que esta intervención basada en las TIC y RA sea de ayuda relevante en estos casos.
- Extender el proceso de intervención psicopedagógica a través de RA utilizando TEARA y el Software Hope a trastornos como el desintegrativo infantil o el generalizado del desarrollo no especificado, el síndrome de Asperger o el de Rett.
- El reto mayor al que se enfrenta las TIC enfocadas a procesos de enseñanza aprendizaje y más aún la RA es conocer si el proceso de intervención pedagógica basado en TEARA se mantiene luego un determinado tiempo (6 meses o 1 año), la intención es verificar si el aprendizaje está almacenado en memoria de largo plazo de la población que participa en las sesiones propuestas.



**Advisor**

Lic. Francisco Javier Diaz

**Scientist Consultant**

Mg. Ivana Harari

**Thesis approval date**

December 6, 2022

**SEDICI**

[Full Text](#)

## **Hope Project: Argument Reality for Teaching-Learning Processes in the Asd Child Population**

**Keywords:** Autism (ASD); User Centered Design (UCD); Hope Project; Software prototype; processes teaching and learning; Augmented Reality (AR); autism spectrum disorder (ASD), Information and Communication Technologies (ICT).

### **Motivation**

Autism defined as a characteristic in which there are variations in distinct aspects, such as a decrease in communication, imagination, and social adaptation. People with Autism Spectrum Disorder (ASD) develop and experience the educational process in particular ways, of the treatments that have used for decades as educational proposals are becoming obsolete over time, there are technological resources based on Augmented Reality (AR) specifically designed to enable accessibility for people with disabilities. This work focuses on these technological resources given their pertinent characteristics that make them highly suitable for use in teaching and learning environments. Therefore, it is intended to deepen this field and be a project that adds and contributes to the research scenario, the objective of this doctoral thesis was the analysis, design, development, implementation and evaluation of the impact of a software prototype that was conceived from an inclusive design centered on the user based on AR and aimed at strengthening the capacities of social learning and cognitive development of the ASD child population. This scientific process covered several years of research and development, being divided into stages, to achieve the proposed final objective. Within academic research, little literature can be seen and recognized, and little incidence related to it, regarding the topic that this paper addresses. Although it is true that AR supports learning, we see that in the ASD child population solutions are not implemented with the frequency and with the desired and required intensity, which justifies further exploration of the advantages of such scenarios. The investigation is not only a theoretical and exploratory investigation. This work applies the knowledge around the use of AR and builds an application

that adds to the range of possibilities of educational agents in these scenarios. Thus, this work tries to delve into a field that, although it is still little explored, does present promising growth rates, so that this work can be seen as an important contribution in this constant sum for the search for better living conditions for the ASD child population. This work yields relevant information that allows, first, to publicize the technological possibilities within educational environments and, second, to positively influence practice and scientific analysis about the possibilities.

### **Thesis contributions**

This research made it possible to analyse, investigate, develop, implement and experiment by creating a prototype that uses AR as a pedagogical tool that complements the educational activities of the ASD child population, which, through an inclusive design and based on a DCU, included criteria of differential quality, usability and functionality compared to the current state of applications aimed at TEA children and based on AR. This is an innovation project to the extent that it integrates branches of knowhow: human-computer interaction, accessibility, user-centered design, intended to offer inclusion for people with disabilities. It conducted a systemic interaction between teaching staff, medical staff, ICT staff, child population and ASD family, in addition to including a multidisciplinary team. The Educational Treatment of Children with ASD mediated through AR called TEARA defined, which designed to strengthen teaching and learning processes by promoting the use of sight, hearing, body perception and movement for children with mild ASD, moderate, or severe can be used by children from 4 years to 16



years. TEARA uses a training system called Software Hope, which teaches artistic expression through dance. This system allows the mobility of the ASD child population, in addition to offering options and configuring the play environment by adding or removing options for its use, it seeks to progressively include new learning, showing greater complexity as time goes by, it can be used in academic center, therapy center, or at home. The design of the software called Hope was not behavioural, it allowed the ASD child population to learn at their own pace; It is an original work, since worldwide software has not been created that uses AR to teach dance to children with ASD, generating new knowledge for us since there is currently no bibliographical source of user-centered inclusive design in addition to contributing with guides to execute usability tests for children with ASD using RA. We evaluate the existing conditions for the implementation of AR in the teaching and learning process of the child population, we design and validate the technological prototype, intended for educational contexts, in addition to defining a Software intervention Methodology designed with AR. We delve into a stage of reflection and discussion about the lived experience based on experimentation in each of the phases. Finally, the use of ICT and AR was carried out in an active, purposeful but not intrusive way, it allowed to demolish certain beliefs where it is indicated that AR technology is addictive and therefore not advisable for the ASD child population, substantial to the learning of processes such as imitation, perception, fine and gross motor skills, as well as visual-motor coordination.

### Future Research Lines

The doctoral research involved years of research on the interaction of the ASD child population and AR technologies. On the other hand, among the lines of research that can be opened from this study, the following can be mentioned:

- Disseminate TEARA as a disruptive way of innovating intervention strategies in teaching-learning processes that are based on integrative experiences to support the teaching and learning processes of children with autism.
- Consider the creation of training programs where TEARA is disseminated in the population (teaching, medical and rehabilitation staff) and the benefits of using the Hope training software for the education of the ASD child population, explaining the hardware and software elements that are required in the intervention.
- Implement recreational spaces based on TEARA, in educational units or care centers so that other types of disabilities can be intervened to strengthen teaching and learning processes through the proposed educational treatment.
- Expanding the study to provide the Hope Training System with new teaching-learning processes in addition to the established ones (imitation, perception, fine and gross motor skills, visual-motor coordination) can be expanded to processes

that include flexibility of thought, memory, and imagination.

- TEARA used to verify teaching-learning processes in ASD adolescents and older adults. Expanding the intervention proposal in children with mild autism, it is possible that this plan and its results are better received, and that this intervention based on ICT and AR is of relevant help in these cases.
- Extend the process of psychopedagogical intervention through AR using TEARA and the Hope Software to disorders such as infantile disintegrative or pervasive developmental disorder not otherwise specified, Asperger syndrome or Rett syndrome.
- The biggest challenge faced by ICT focused on teaching-learning processes and even more AR is to know if the TEARA-based pedagogical intervention process maintained after a certain time (6 months or 1 year), the intention is to verify if learning stored in long-term memory of the population that participates in the proposed sessions.

## DOCTORADO EN CIENCIAS INFORMÁTICAS



**Dr. Ignacio Martin Gallardo Urbini**

e-mail

ignaciomgu@gmail.com

español

**Directora**

Dra. Patricia Bazán

**Asesores Científicos**

Mg. Paula Venosa

Mg. Nicolás del Río

**Fecha de aprobación**

13 de diciembre de 2022

**SEDICI**

[Tesis Completa](#)

## **Estrategia de Ciberseguridad distribuida, aplicando el concepto de Operación de Inteligencia**

**Palabras clave:** Ciberseguridad; Ciberdefensa; Operaciones de inteligencia; Inteligencia de datos; Aprendizaje automático; Multi vendedor; Espías; Inteligencia de amenazas

### **Motivación**

Las soluciones de seguridad tradicionales están orientadas principalmente en proteger el perímetro de interés y, por lo tanto, se centran principalmente en las amenazas externas. Sin embargo, estos están en constante evolución, lo que requiere que aquellos que desean seguir siendo resistentes en sus operaciones se mantengan informados y un paso por delante de los atacantes. Para la definición de una estrategia defensiva de ciberseguridad se pueden utilizar las mismas variables que se toman en cuenta en la doctrina de inteligencia aplicada a la seguridad nacional, donde se presentan elementos de agresión similares a los analizados en un ciberataque: sabotaje, acoso a la víctima en su suelo propio, uso de destacamentos irregulares con ataques rápidos y sorpresivos, secretismo, gran movilidad, bloqueos temporales de los canales básicos de comunicación y abastecimiento, y secuestro / robo de bienes. Ante este nuevo contexto de ciberamenazas avanzadas, en el que se involucran grupos criminales y hacktivistas con intereses políticos y económicos, surge la motivación de iniciar esta línea de investigación con el fin de llevar a cabo el desarrollo de la inteligencia o estrategia de ciberinteligencia como elemento clave. para reforzar la estrategia de seguridad de la información.

### **Aportes de la tesis**

El marco propuesto en esta tesis incluye tácticas, estrategias, y procedimientos aplicados en operaciones de inteligencia incluidos en la propia doctrina de inteligencia nacional, en español y para uso público, con un protocolo de comunicación abierto para compartir datos y el conocimiento aprendido para ser utilizado o consumido por con agentes externos, integrando recolectores de datos, módulos de detección de anomalías adaptables y un marco de referencia para adelantarse al enemigo y así poder tomar una acción disuasoria, ofensiva o defensiva.

### **Líneas de I/D futuras**

Si bien se ha demostrado una solución eficaz, eficiente y funcional; la creación de un producto de software sólido, con métricas de tiempos de procesamientos y consumo de recursos, queda fuera del alcance de esta tesis. Por lo tanto, se propone como trabajo a futuro seguir mejorando esta solución, abriendo el código a la comunidad para que puedan extender a su vez con nuevas funcionalidades y avances.

Una extensión funcional posible podría ser el soporte de más módulos de predicción, por ejemplo, entrenar nuevos modelos o incluso utilizar otros ya existentes, como los creado por Paula Venosa en su tesis de maestría: "Detección de ataques de seguridad en redes usando técnicas de ensembling". Esto podría darle una gran funcionalidad a la arquitectura propuesta en esta tesis para detectar ataques a nivel de red.

Si bien no es incumbencia de la tesis, ya que el Ciclo de Vida de la Inteligencia termina en la Fase de Difusión, sería interesante poder integrar los Retro-Alimentadores y Alertadores con más interfaces externas, como por ejemplo, más variedad de Firewalls y otros servicios como ser llamadas telefónicas y montar toda esta la infraestructura en servidores y máquinas más potentes para mejorar el rendimiento.

Otro aspecto a destacar, es que en este framework propuesto se trató en la medida de lo posible, de no acotarlo o acoplarlo a tecnologías específicas, sin embargo, una perspectiva de interés a corto-mediano plazo, es integrar a esta solución más herramientas de big data, como ser la suite Hadoop y Spark.

Finalmente, se propone también como trabajo a futuro, la creación de librerías Proveedor de Transacciones en distintos lenguajes de programación, para poder integrarlos a más cantidad de aplicaciones webs o incluso dispositivos en general.



english

**Advisor**  
Dra. Patricia Bazán  
**Scientists Consultants**  
Mg. Paula Venosa  
Mg. Nicolás del Río

**Thesis approval date**  
December 13, 2022

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

## **Distributed Cybersecurity Strategy, applying the Intelligence Operation theory**

**Keywords:** Cyber-security; Cyber-defence; Intelligence Operations; Data Intelligence; Machine Learning; Multivendor; Spies; Threats Intelligence.

### **Motivation**

Traditional security solutions focus primarily on protecting the perimeter of interest, thus focusing primarily on external threats. Yet these are constantly evolving, requiring those who wish to remain resilient in their operations to stay informed and one step ahead of attackers. For the definition of a defensive cybersecurity strategy, the same variables that are taken into account in the intelligence doctrine applied to national security can be used, where elements of aggression similar to those analysed in a cyberattack are presented: sabotage, harassment of the victim in his land, use of irregular detachments with rapid and surprise attacks, secrecy, great mobility, temporary blockages of the basic channels of communication and supplies, and kidnapping/ theft of assets. Faced with this new context of advanced cyber threats, in which criminal and hacktivist groups with political and economic interests are involved, the motivation arises to start this line of investigation to carry out the development of intelligence or cyber intelligence strategy as an element key to reinforcing the information security strategy.

### **Thesis contributions**

The framework proposed in this thesis includes tactics and strategies, and procedures applied in intelligence operations included in the national intelligence doctrine itself, in Spanish and for public use, with an open communication protocol for sharing the learned knowledge with external vendors to be used or consuming data from them, integrating data collectors, adaptable anomaly detection modules and a frame of reference to get ahead of the enemy and thus be able to take a dissuasive, offensive or defensive action.

### **Future Research Lines**

Although an effective, efficient and functional solution has been demonstrated, creating a solid software product, including processing times and resource consumption metrics, is out of this thesis's scope. Therefore, it is proposed as future work to continue improving this solution, opening the code to the community to extend it with new features and improvements. A functional extension could support more prediction modules, for example, training new models or even using existing ones, such as those created by Paula Venosa in her master's thesis: "Detection of security attacks in networks using ensemble techniques". This could bring great functionality to the architecture proposed in this thesis to detect attacks at the network level. Notwithstanding, it is not the responsibility of the thesis since the Intelligence Life Cycle ends in the Diffusion Phase, it would be interesting to be able to integrate Feedback, Feeders, and Whistleblowers with more external interfaces, such as, for example, a greater variety of Firewalls and others services such as telephone calls and even hosting all this infrastructure over more powerful servers and machines to improve performance. Another aspect to highlight is that this proposed framework will be analysed to the extent possible, not limiting it or coupling it to specific technologies; however, a perspective of interest in the short-medium term is to integrate more tools of big data, such as the Hadoop and Spark suite. Finally, it is also proposed as future work the creation of Transaction Provider libraries in different programming languages to integrate them into web applications or even devices.



**Director**  
Ing. Armando De Giusti

**Codirector**  
Dr. Iván Danilo García Santillán

**Fecha de aprobación**  
28 de diciembre de 2022

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

## Algoritmos para Agricultura de Precisión utilizando Computación de Alto Rendimiento

**Palabras clave:** Agricultura de precisión; Algoritmos paralelos; Análisis de imágenes; Detección de líneas de cultivo; Detección de malezas; HPC; Matlab; UAV

### Motivación

La agricultura de precisión (AP) automatiza actividades mediante la recopilación y análisis de datos agrícolas para la toma de decisiones con fines de mejorar la producción agrícola. En AP el procesamiento de imágenes digitales agrícolas implica operaciones de procesamiento para obtener diferentes funcionalidades en la sistematización de tareas agrícolas necesitando varios recursos de cómputo, principalmente de tiempo, memoria y procesador.

El cambio tecnológico, fundamentalmente a partir de los procesadores con varios núcleos (multicore) y otras arquitecturas heterogéneas actuales (gpu, fpga, vtu) ofrecen la posibilidad de utilizar nuevos paradigmas de hardware y software (técnicas de programación), en los cuales coexisten esquemas de memoria compartida, siendo una buena alternativa para alcanzar un alto rendimiento en determinadas aplicaciones informáticas. Los drones (UAV Unmanned Aerial Vehicle) pueden recorrer campos de cultivo extensos en periodos cortos de tiempo para capturar imágenes y posteriormente ser utilizadas en aplicaciones agrícolas con múltiples funciones como: mapeo de campos, monitoreo de cultivos, detección de enfermedades, detección de plagas, detección de líneas de cultivo calidad de suelo, aplicación de fungicidas y plaguicidas, entre otros. Las técnicas de análisis de imágenes y visión por computador son utilizadas en la AP para extraer datos importantes almacenados como píxeles; por ello requieren un alto costo computacional en la automatización de actividades agrícolas para mejorar aspectos como la precisión, cobertura y tiempos de ejecución. La detección temprana de malas hierbas en cualquier tipo de cultivo puede evitar pérdidas en las cosechas; generando mayor productividad a los agricultores. Para solventar los retos anteriormente

mencionados se propone un objetivo general para desarrollar algoritmos de visión por computadora utilizando técnicas de programación de alto rendimiento para detección automática de líneas de cultivo y malas hierbas en campos de maíz, estudiando su optimización en cuanto a la arquitectura de soporte y el tiempo de respuesta. Los objetivos específicos que complementan el general son los siguientes:

- Describir el estado del arte respecto a las arquitecturas paralelas, técnicas de programación de alto rendimiento y algoritmos de visión por computador para detección automática de líneas de cultivo y malas hierbas.
- Diseñar algoritmos de visión por computador para la detección de líneas de cultivo y malas hierbas en campos de maíz utilizando técnicas de programación de alto rendimiento.
- Implementar algoritmos de visión por computador para la detección de líneas de cultivo y malas hierbas en arquitecturas multiprocesador utilizando Matlab.
- Evaluar el rendimiento de los métodos propuestos y versiones secuenciales originales utilizando métricas cuantitativas para arquitecturas multiprocesador.

### Aportes de la tesis

El aporte general de la propuesta está enfocado en la investigación y desarrollo de algoritmos de visión por computador utilizando técnicas de programación de alto rendimiento en plataformas de hardware paralelas mediante procesamiento de imágenes adquiridas con drones para detección automática de líneas de cultivo y malas hierbas en campos de maíz.

Teniendo en cuenta las pérdidas económicas en la agricultura cada vez mayores debido a las malezas en cultivos de

maíz, se propone un algoritmo paralelo escalable aplicando la metodología de Foster's capaz de identificar líneas de cultivo y discriminar malezas para implementar en arquitecturas multiprocesador con memoria compartida, con el objetivo de mejorar el rendimiento de los algoritmos reduciendo los tiempos de ejecución en el procesamiento de imágenes digitales agrícolas.

El algoritmo se ejecuta de manera paralela iniciando con 4 particiones principales (segmentación, detección de líneas de cultivo, exclusión de cultivo, discriminación de maleza), en cada partición existen diferentes tareas relacionadas con el análisis de imágenes para luego ser enviadas al procesador con varios núcleos (multicore). Se utilizó Matlab como lenguaje de programación y herramientas orientadas al paralelismo local para procesar imágenes adquiridas mediante un dron DJI Mavic 2 Pro con una resolución de 5472× 3648 a alturas de 5, 10 y 15 metros. Los resultados obtenidos demuestran que se puede identificar como mínimo líneas de cultivo (85%) y máximo de maleza (93.28%) del total de vegetación en imágenes de la cuarta semana de seguimiento al cultivo a 15 metros de altura. Igualmente, con el algoritmo propuesto los tiempos de procesamiento evaluados en promedio alcanzan un rango entre 4.57 y 5.24 segundos con imágenes que alcanzan una extensión máxima de 114 metros cuadrados. Adicionalmente se puede verificar que el algoritmo es escalable porque reduce el tiempo de ejecución a medida que el número de procesadores crece.

### **Líneas de I/D futuras**

- Investigación en temas de cómputo paralelo y distribuido de alto desempeño, en lo referido a los fundamentos y a la construcción y evaluación de las aplicaciones
- Construcción, evaluación y optimización de soluciones utilizando algoritmos concurrentes, paralelos y distribuidos sobre diferentes plataformas de software y arquitecturas con múltiples procesadores:
- Análisis y desarrollo de ambientes para la enseñanza de programación concurrente y paralela.
- Desarrollo de aplicaciones de software para diversos tipos de cultivos.



**Advisor**  
Ing. Armando De Giusti

**Codirector**  
Dr. Iván Danilo García Santillán

**Thesis approval date**  
December 28, 2022

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

## Algorithms for Precision Agriculture using High Performance Computing

**Keywords:** Precision agriculture; Parallel algorithms; Image analysis; Crop row detection; Weed detection; HPC; Matlab; UAV;

### Motivation

Precision agriculture (PA) automates activities by collecting and analyzing agricultural data for decision-making to improve agricultural production. In AP, processing agricultural digital images involves processing operations to obtain different functionalities in the systematization of farming tasks, requiring various computing resources, mainly time, memory, and processor.

The technological change, fundamentally from the processors with several cores (multicore) and other current heterogeneous architectures (GPU, FPGA, VTU) offers the possibility of using new hardware and software paradigms (programming techniques), in which shared memory schemes coexist, being a good alternative to achieve high performance in certain computer applications. Drones (UAV- Unmanned Aerial Vehicles) can travel extensive fields in short periods to capture images and subsequently be used in agricultural applications with multiple functions such as field mapping, crop monitoring, disease detection, pest detection, detection of soil quality crop rows, application of fungicides and pesticides, among others.

Image analysis and computer vision techniques are used in THE AP to extract important data stored as pixels; therefore, they require a high computational cost to automate agricultural activities to improve accuracy, coverage, and execution times. Early detection of herbs in any crop can prevent crop losses, generating greater productivity for farmers. In order to solve the aforementioned challenges, a general objective is proposed to develop computer vision algorithms using high-performance programming techniques for automatically detecting crop rows and weeds in corn fields, studying their optimization in support architecture and response time. The

specific objectives that complement the general one is the following:

- Describe state of the art regarding parallel architectures, high-performance programming techniques, and computer vision algorithms for automatically detecting crop rows and weeds.
- Design computer vision algorithms for the detection of crop rows and weeds in cornfields using high-yield programming techniques.
- Implement computer vision algorithms for the detection of crop rows and weeds in multiprocessor architectures using Matlab.
- Evaluate the performance of proposed methods and original sequential versions using quantitative metrics for multiprocessor architectures.

### Thesis contributions

The general contribution of the proposal is focused on the research and development of computer vision algorithms using high-performance programming techniques in parallel hardware platforms through image processing acquired with drones for the automatic detection of crop rows and weeds in corn fields.

Considering the increasing economic losses in agriculture due to weeds in corn crops, a scalable parallel algorithm is proposed applying Foster's methodology capable of identifying crop rows and discriminating weeds to implement in multiprocessor architectures with shared memory to improve the performance of the algorithms by reducing the execution times in the processing of agricultural digital images.

The algorithm runs in parallel, starting with four main partitions (segmentation, detection of crop rows, exclusion of culture, and weed discrimination); in each partition, there

are different tasks related to the analysis of images to then be sent to the processor with several cores a programming language and tools oriented to local parallelism to process images acquired by a DJI Mavic 2 Pro drone with a resolution of 5472× 3648 at heights of 5, 10, and 15 meters. The results obtained show that it is possible to identify at least crop rows (85%) and maximum weeds (93.28%) of the total vegetation in images of the fourth week of monitoring the crop at 15 meters high. Likewise, with the proposed algorithm, the processing times evaluated on average reach a range between 4.57 and 5.24 seconds with images that reach a maximum extension of 114 square meters. Additionally, it can be verified that the algorithm is scalable because it reduces the execution time as the number of processors grows.

### **Future Research Lines**

- Research on high-performance parallel and distributed computing issues, in terms of fundamentals and the construction and evaluation of applications
- Construction, evaluation, and optimization of solutions using concurrent, parallel, and distributed algorithms on different software platforms and multi-processor architectures:
- Analysis and development of environments for the teaching of concurrent and parallel programming. Development of software applications for various types of crops





02

MAESTRÍAS

TECNOLOGÍA INFORMÁTICA  
APLICADA EN EDUCACIÓN

INGENIERÍA DE SOFTWARE

REDES DE DATOS



español

**Director**  
Dr. Matías Urbieto  
**Fecha de aprobación**  
9 de marzo de 2022  
**SEDICI**  
[Tesis Completa](#)

## **Arquitectura de Software en Wallet de código abierto para privacy coin en dispositivos móviles. El caso de estudio de Zcash**

**Palabras clave:** Zcash; Privacy coin; Aplicaciones Móviles; Cliente liviano; Arquitectura; Criptomoneda

### **Motivación**

Desde su surgimiento en 2009, el uso de criptomonedas ha estado en constante crecimiento en términos de cuota de mercado y adopción. Este “boom” está a la vista de todos, y al igual que la gran mayoría de las transacciones del mundo de las finanzas descentralizadas, es de completo estado público. Pese al uso de criptografía de avanzada, la privacidad en el “mundo cripto” es relativamente baja, con excepciones: Las Privacy Coins o Monedas con Mejoramiento del Anonimato (Anonymity Enhancement Coins o AEC). Este trabajo toma la idea de la privacidad como Derecho Humano y se centra en dilucidar los requerimientos para el desarrollo de billeteras electrónicas móviles para AECs, analizando las criptomonedas Monero y Zcash, tomando como caso de estudio esta última.

### **Aportes de la tesis**

Tiene como aporte una lista de requerimientos funcionales y no funcionales y proponiendo una arquitectura de referencia para cumplir con los mismos de forma abstracta, e identifica áreas a profundizar en materia de revisión sistemática de la literatura, privacidad y seguridad.

### **Líneas de I/D futuras**

- Ampliación de la metodología de Revisión Sistemática de la Literatura para contemplar y normalizar las fuentes de consulta al ámbito de las criptomonedas
- Marco de Métricas de Referencia para la evaluación del estado de Proyectos de Criptomonedas
- Evaluación inclusión del protocolo ‘FlyClient’ en la presente propuesta
- Evaluación de esquemas sociales para la preservación y recuperación de claves privadas
- Estrategias de protección de la privacidad del tráfico de red.



english

**Advisor**  
Dr. Matías Urbieta  
**Thesis approval date**  
March 9, 2022  
**SEDICI**  
[Full Text](#)

## **Software Architecture for Open Source Mobile Privacy Coin Mobile Wallet applications. The Zcash Case Study**

**Keywords:** Zcash; Privacy coin; Mobile; Light client; Architecture; Cryptocurrency

### **Motivation**

Since their appearance in 2009, the use of cryptocurrencies has been growing constantly in terms of market cap and adoption. This boom is publicly visible as well as the grand majority of the decentralized finance transactions. Despite the use of advanced cryptography, privacy in the “crypto world” is relatively low, with certain exceptions: Privacy Coins (or Anonymity Enhanced Coins AEC). This work takes the idea of Privacy as a Human Right and focuses on eliciting the requirements for developing mobile wallets for AECs, analyzing the cryptocurrencies Monero and primarily Zcash, taking the latter as study case.

### **Thesis contributions**

Its contributions are: a list of functional and non-functional requirements to develop a privacy coin light client, a reference Architecture that addresses these requirements in an abstract manner and finally a list of future work related to the fields of Systematic Literature Review, Privacy and Security.

### **Future Research Lines**

- Systematic Literature Review Methodology Enhancement Applied to the Cryptocurrency space.
- Metrics to evaluate the viability of Open Source Cryptocurrency projects.
- Adding FlyClient protocol to the present proposal.
- Analysis of social seed phrase recovery on mobile wallets.
- Network traffic Privacy strategies for light client information retrieval.



Director  
Dr. Matías Urbietta  
Fecha de aprobación  
14 de marzo de 2022

## Modelos de Predicción Avanzados para el cálculo de Reservas en la Industria Aseguradora

**Palabras clave:** MachineLearning; Reservas; Aseguradora; Seguro; Predictivo; Lineal; ARIMA; LSTM

### Motivación

La industria del seguro tiene una necesidad urgente de innovación luego del arribo del concepto de “Insurtech”, que propone un mercado más competitivo basado en la innovación y en la aplicación de tecnología al negocio (Alarcon, 2018). Pero por otro lado se trata de una actividad muy regulada por los organismos gubernamentales, que establecen un marco común de aplicación de procedimientos y técnicas que las aseguradoras deben respetar.

En esa puja entre innovación y regulación, está el concepto de las Reservas, que representan la provisión económica que las empresas aseguradoras deben disponer para hacer frente al pago de los posibles siniestros que podrían ocurrir en el futuro, como consecuencia directa de las coberturas que les brindan a sus asegurados (NietoDeAlba, 1964).

Si bien se trata de un tema definido por ley, representa uno de los indicadores más importantes a tener en cuenta en la toma de decisiones porque establece el límite de la libre disponibilidad de fondos que la compañía puede invertir para obtener resultados financieros (LopezBriega, 2016). No preocupa tanto si en un período un rubro o producto ofrecido por una aseguradora es deficitario si se tiene en cuenta sus resultados técnicos, porque la ganancia se puede generar a partir de los resultados financieros producto de invertir el dinero en el tiempo transcurrido entre que se cobra el seguro y se paga el siniestro. En esa actividad cuando mejor sea la predicción de las reservas, mayor será la renta financiera obtenida de las disponibilidades monetarias, mayor será el ahorro de costos directos, y se tendrían más oportunidades en la aplicación de precios personalizados que permitan capturar clientes.

El objetivo general de éste trabajo es analizar el funcionamiento de diferentes modelos de predicción avanzados, basados en técnicas de Machine Learning, que puedan estimar las reservas futuras con antelación a la ejecución de

los tradicionales procedimientos de cálculo usados en la industria de acuerdo a lo definido por la ley.

#### Objetivos específicos

- Relevar los tipos de reservas en la industria aseguradora.
- Relevar su procedimiento de cálculo tradicional.
- Definir las técnicas de Machine Learning y algoritmos de predicción a utilizar.
- Recolectar y preparar sets de datos.
- Ejecutar experimentos y obtener el mejor modelo de predicción posible de cada técnica.
- Analizar resultados y comparar métricas.
- Concluir acerca del funcionamiento de cada mecanismo.

#### Aportes de la tesis

A partir de un muestreo de datos de reserva real de una compañía aseguradora, se exploraron diferentes modelos de predicción de reserva tanto de seguros generales como de ramos en particular, desagregadas a nivel de contratos individuales y acumuladas mensualmente como proyecciones en una serie de tiempo.

Los modelos se entrenaron en diferentes monedas, haciendo especial análisis en el impacto de procesos inflacionarios.

Se utilizaron las técnicas de Regresión Lineal, ARIMA y Redes Neuronales LSTM, y un proceso metodológico adaptado a Machine Learning basado en el estándar CRISP-DM. Se pudieron obtener las siguientes conclusiones:

- Fue significativo el tiempo y la complejidad de los procedimientos administrativos para conseguir las autorizaciones necesarias para poder acceder a los datos de la compañía de seguros. Se trata de una empresa grande con presencia nacional e internacional, con fuertes estándares de seguridad y auditorías que regulan en acceso a la información de negocio.
- Fue significativo el tiempo incurrido en el entendimiento de

los modelos de datos de los cuales obtener la información, dada la heterogeneidad de sistemas “legados” y nuevos, con documentación incompleta en algunos casos, que fue completada a través de entrevistas con personal idóneo en IT.

- Los dos puntos anteriores sumado al volumen de los datos, hicieron que el 60% del tiempo total del proceso de Machine Learning se incurra en el acceso, entendimiento y preparación de los sets de datos.

- Se realizó el entrenamiento, validación y testing de modelos lineales que predigan la reserva de una póliza individual de seguros automotor, utilizando la prima y capital asegurado como las características de entrada del modelo, pero no se consiguieron niveles de error satisfactorios que cumplan los criterios de evaluación definidos.

- Se realizó el entrenamiento de modelos lineales y técnicas especializadas en series temporales (ARIMA y Redes LSTM) que predigan la reserva acumulada en forma mensual, tanto a nivel general como desagregada en los ramos de Automotores, Vida colectivo y Combinado Familiar. Si bien la cantidad de datos era bajo (un valor por mes durante 3 años), todos los modelos cumplieron el criterio de aceptación definido en un 20% de error MAPE.

- Si bien los modelos lineales se ajustaron efectivamente a la tendencia de los datos de la serie de tiempo, no pudieron manejar los ciclos estacionales por lo que el error fue significativamente más altos en comparación con los algoritmos especializados en series temporales que sí manejan tanto tendencia como estacionalidad.

- Los modelos entrenados con datos expresados en dólares estadounidenses (convertidos utilizando el cambio oficial de cada mes), tuvieron menor error en comparación al entrenamiento en pesos argentinos, ya que tienen menos influencia del contexto devaluatorio e inflacionario que caracterizó a nuestro país los últimos años.

- Si bien todos los modelos cumplieron los criterios de aceptación en sus predicciones para los meses en los que fueron entrenados, el error aumentó cuando realizaron predicciones a futuro. Ésta situación se vio agravada por un contexto especial, donde ocurrieron cambios normativos a nivel nacional para algunos ramos, y situaciones excepcionales como la pandemia mundial de COVID-19.

- El error de los modelos mejoró y volvió a cumplir los criterios de aceptación al someterlos a un proceso de reentrenamiento mensual, sumando mes a mes los nuevos valores de reservas calculadas por los procesos tradicionales al set de datos de entrenamiento de los modelos.

- El error más bajo en las predicciones lo obtuvo la técnica ARIMA entrenada sobre datos en dólares para los casos en donde no hubo cambios bruscos en los datos de un mes a otro, mientras que las redes neuronales LSTM, entrenadas también con datos expresados en dólares, tuvieron mejor desempeño al momento de realizar predicciones en los ramos que sufrieron cambios bruscos de tendencia.

## Líneas de I/D futuras

La industria del seguro genera oportunidades para el desarrollo de técnicas de Machine Learning. En ésta investigación se trabajó sobre reservas, pero existen otras aplicaciones que se pueden explorar como siniestralidad, producción, fidelización de clientes, detección de morosidad y clientes deficitarios, reconocimiento de patrones en imágenes de siniestros agrícolas y patrimoniales, zonas de riesgos, etc.. Del presente trabajo se desprenden las siguientes líneas concretas de investigación:

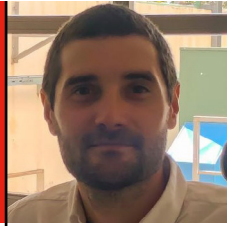
- Trabajar con otros modelos predictivos de reservas en pólizas individuales que mejoren los indicadores de error y puedan cumplir con el criterio de aceptación definido.

- Analizar la aplicación de ésta técnica para predecir otros indicadores de importancia en el mundo del seguro como siniestralidad y producción.

- Ampliar los algoritmos candidatos a ser utilizados en las predicciones y análisis de error.

- Profundizar en la parametrización de los modelo ARIMA, analizando la posibilidad de prevenir los puntos extremos en las predicciones en el mes siguiente a un cambio brusco en la tendencia de los datos.

- Analizar la posibilidad de reducir la complejidad de los modelos de redes LSMT sin afectar los niveles de error conseguidos.



**Advisor**  
Dr. Matías Urbieta  
**Thesis approval date**  
March 14, 2022

## **Advanced Prediction Models for Reserves in the Insurance Industry**

**Keywords:** MachineLearning; Reserves; Insurance; Predictive; Lineal; ARIMA; LSTM

### **Motivation**

The insurance industry has an urgent need for innovation after the arrival of the “Insurtech” concept, which proposes a much more competitive business based on innovation and the application of technology (Alarcon, 2018). However, it’s an activity highly regulated by government agencies, which establish common procedures and techniques that insurance companies must respect.

In this context between innovation and regulation, exist the concept of Reserves, which represent the economic provision that the companies must have to cover with the payment of possible claims that could occur in the future, as a direct consequence of the coverage that they provide to their clients (NietoDeAlba, 1964).

Although it is a concept defined by law, it represents one of the most important indicators in the organization’s decisions because it establishes the limit on the free availability of economic resources that the company can invest for financial results (LopezBriega, 2016). It is not so much concern if in a period a product offered by the insurer company has a deficit in its technical results, because the rent can be generated from the financial results of investing the money in the time elapsed between the collection of the insurance and the payment of the claim. In this activity, a good prediction of the reserves will generate improvements in the financial income that can be obtained from the monetary availability, direct cost savings, and better prices to capture new customers.

The general objective of this work is to analyze the operation of different advanced prediction models, based on machine learning techniques that can estimate future reserves in advance of the execution of traditional calculation procedures defined by law.

#### **The specific objectives are:**

- Exploration of the types of reserves on the insurance industry.

- Exploration of the traditional calculation procedure.
- Definition of the Machine Learning techniques and prediction algorithms to use.
- Collect and prepare data sets.
- Run experiments and obtain the best possible prediction model for each technique.
- Analyze results and compare metrics.
- Obtain conclusions about the workings of each mechanism.

### **Thesis contributions**

Based on a sample of real reserve data from an insurance company, different reserve prediction models were explored, both by general insurance and by branch of business in particular, disaggregated at the level of individual contracts and accumulated monthly as projections in a time series.

The models were trained in different currencies, making special analysis on the impact of inflationary processes.

The techniques used were Linear Regression, ARIMA and LSTM neural networks, and a methodological process adapted to Machine Learning based on the CRISP-DM standard.

The following conclusions were obtained:

- The time and complexity of the administrative procedures to obtain the necessary authorizations to access the insurance company data was significant. It is a large company with a national and international presence, with strong security standards and audits that regulate access to business information.
- The time invested in understanding the data models from which to obtain the information was significant, due to the heterogeneity of “legacy” and new systems, with incomplete documentation in some cases, which was completed with interviews along IT professionals.

- The two previous items in addition to the volume of the data, made it so that 60 % of the total time of the Machine Learning process to be consumed in accessing, understanding and preparing the data to was used for the work.

- The process of training, validation and testing of linear models that predict the reservation of an individual automobile insurance policy was executed, using the premium and the insured capital as the input characteristics of the model, but satisfactory levels of error that met the criteria of defined evaluation were not achieved.

- The process of training was executed on linear models and specialized techniques in time series (ARIMA and LSTM Networks) that predict the accumulated reserve on a monthly basis, both at a general level and disaggregated in the Automotive, Collective Life and Combined Family branch.

Although the size of the dataset was low (one value per month for 3 years), all models met the acceptance criteria defined in an 20 % of the MAPE error.

- Although the linear models did adjust effectively to the trend of the time series, they could not handle the seasonal cycles so the error were significantly higher compared to the specialized time series algorithms that do handle both trend as seasonality.

- The models trained with data expressed in US dollars (converted using the official exchange of each month), had lower error compared to the training in Argentine currency, because they are less influenced by the devaluation and inflationary context than characterized our country in recent years.

- Although all the models met the acceptance criteria in their predictions for the months in which they were trained, the error increased when they made forward predictions. This situation was aggravated by a special context, where regulatory changes occurred at the national level for some branches, and exceptional situations such as the global COVID-19 pandemic.

- The error of the models improved and returned to meet the acceptance criteria by subjecting them to a monthly retraining process, adding month by month the new reserve values calculated by the traditional processes to the training data set of the models.

- The lowest error in the predictions was obtained by the ARIMA technique trained with data expressed in US dollars, for cases where there was no sudden changes in the data from one month to another.

While the LSTM neural networks, also trained with data expressed in US dollars, performed better when making predictions in the branches that suffered sudden trend changes.

The insurance industry generates opportunities for the development of Machine Learning techniques. In this investigation we worked in reserves, but there are other application that can be explored such as claims, production, customer loyalty, detection of delinquency and deficit customers, recognition of patterns in images of agricultural and property claims, risk zones, etc..

The following specific lines of research emerge from this work:

- Work with other predictive models of reserves in individual policies that improve the error and can meet the defined acceptance criteria.

- Analyze the application of this technique to predict other important indicators in the insurance industry such as claims and production.

- Expand the candidate algorithms used in predictions and error analysis.

- Review the parameterization of the ARIMA models, analyzing the possibility of preventing extreme points in the predictions of the month after a sudden trend change of the data.

- Analyze the possibility of reducing the complexity of LSMT network models without affecting the error.

## Future Research Lines





español

**Director**  
Dr. Alejandro Fernández

**Codirectors**  
MSc. Diego Alberto  
Aracena Pizarro

**Fecha de aprobación**  
20 de abril de 2022

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

## **Detección y Clasificación de Enfermedades en el Tomate Mediante Deep Learning y Computer Vision**

**Palabras clave:** Enfermedades en la planta de tomate; Redes neuronales convolucionales; Detección de objetos; Clasificación de imágenes; Aprendizaje profundo; Visión por computadora.

### **Motivación**

El aprendizaje profundo es una técnica emergente de inteligencia artificial (IA) que utiliza estructuras de análisis sofisticadas llamadas redes neuronales para realizar asociaciones precisas dentro de un conjunto de datos. En particular, los sistemas de aprendizaje profundo pueden aprender procesando datos en bruto sin reglas codificadas por humanos o conocimiento de dominio. Estos sistemas son particularmente expertos en el lenguaje y la clasificación de imágenes, donde un patrón puede representar una idea abstracta como sentimiento, intención o incluso el concepto general de cómo se ve un gato o un perro. Estos sistemas también son excelentes para hacer predicciones, como el comportamiento de los clientes o las previsiones meteorológicas a largo plazo. ¡También hay un increíble potencial para el análisis de imágenes de frutas y hortalizas, para diagnosticar enfermedades y plagas, facilitando y agilizando la detección y diagnóstico de plagas y/o enfermedades presente en las plantas para luego realizar el tratamiento respectivo.

El objetivo general de este trabajo es realizar investigación que se constituya en un aporte creativo a nivel nacional al aplicar el aprendizaje profundo a problemas de visión de computadora tales como detección y clasificación específicamente en las enfermedades del tomate mediante el procesamiento de imágenes digitales.

### **Aportes de la tesis**

Los aportes de la tesis están basados en dos pilares; la primera de ella en la contribución en el desarrollo de una metodología de trabajo para la detección y clasificación de enfer-

medades mediante el procesamiento de imágenes digitales. La segunda mediante la automatización en tiempo real de los procesos de detección, clasificación de enfermedades, advirtiendo sobre una posible amenaza para el cultivo y su posterior tratamiento mediante el procesamiento de imágenes digitales se logra reducir el esfuerzo requerido para el agricultor productor de tomates, ahorrándose el trabajo de desplazarse geográficamente y el tener que ubicar un profesional agrónomo para ese cometido, especialmente cuando el cultivo puede estar en un lugar remoto, además del diagnóstico de la enfermedad prescribir el uso de pesticidas apropiados a ser utilizados por los agricultores, limitando drásticamente la adquisición descontrolada plaguicidas que conducen a un uso excesivo e inadecuado con efectos consiguientemente negativos para el medio ambiente.

### **Líneas de I/D futuras**

A partir del trabajo realizado se puede añadir otras enfermedades como también incorporar la detección y clasificación de plagas que se presentan tanto en la planta como en el fruto del tomate. Además se puede escalar a otros tipos de hortalizas y frutos en base a la metodología de trabajo que se desarrolló para la detección y clasificación de enfermedades y plagas con su respectivo tratamiento.





english

**Advisor**  
Dr. Alejandro Fernández

**Codirectors**  
MSc. Diego Alberto  
Aracena Pizarro

**Thesis approval date**  
April 20, 2022

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

## Detection and Classification of Diseases in Tomato Using Deep Learning and Computer Vision

**Keywords:** Tomato plant diseases; Convolutional neural networks; Object detection; Image classification; Deep learning; Computer vision

### Motivation

Deep learning is an emerging artificial intelligence (AI) technique that uses sophisticated analysis structures called neural networks to make precise associations within a data set. In particular, deep learning systems can learn by processing raw data without human-coded rules or domain knowledge. These systems are particularly adept at language and image classification, where a pattern can represent an abstract idea such as feeling, intention, or even the general concept of what a cat or dog looks like. These systems are also great for making predictions, such as customer behavior or long-term weather forecasts. There is also an incredible potential for the analysis of images of fruits and vegetables, to diagnose diseases and pests, facilitating and speeding up the detection and diagnosis of pests and / or diseases present in plants and then carry out the respective treatment.

The general objective of this work is to carry out research that constitutes a creative nationwide contribution level by applying deep learning to computer vision problems such as detection and classification specifically in tomato diseases through digital image processing.

### Thesis contributions

The contributions of the thesis are based on two pillars; the first one in the contribution in the development of a work methodology for the detection and classification of diseases through digital image processing. The second through the automation in real time of the detection processes, classification of diseases and warning about a possible threat to the crop and its subsequent

treatment through digital image processing, it is possible to reduce the effort required for the tomato producer, saving time by avoiding the traveling hassle and having to locate an agronomist for that task, especially when the cropfield may be in a remote place, in addition to diagnosing the disease, prescribing the use of appropriate pesticides to be used by farmers, drastically limiting the uncontrolled acquisition of pesticides leading to excessive and inappropriate use with consequent negative effects on the environment.

### Future Research Lines

From the work carried out, other diseases can be added as well as the detection and classification of pests that occur both in the plant and the tomato fruit. In addition, it can be scaled to other types of vegetables and fruits based on the work methodology that was developed for the detection and classification of diseases and pests with their respective treatment.



Director  
Mg. Alejandro González

Asesora Científica  
Dra. Alcira Vallejo

Fecha de aprobación  
16 de marzo de 2022

SEDICI  
[Tesis Completa](#)

## Estrategias de utilización de un simulador de entrevistas de trabajo para adultos hipoacúsicos con ayudas auditivas

**Palabras clave:** Tecnología aplicada en Educación; Simuladores educativos; Hipoacusia; Capacitación en línea; Entrevista Laboral

### Motivación

La presente tesis desarrolla un prototipo de simulador en línea como herramienta de formación inclusiva para adultos postlocutivos.

#### Objetivos específicos

(seguridad en comunicación- inteligencia emocional)

- Analizar las fortalezas y debilidades en la adquisición de nuevos conocimientos por parte de adultos hipoacúsicos postlocutivos.
- Estudiar el concepto de simulador en línea y sus fundamentaciones en el contexto educativo y de formación.
- Desarrollar un prototipo de simulador que sirva de entrenamiento para entrevistas laborales.
- Evaluar el grado de efectividad en estrategias que favorecen la atención, la memoria, la estrategia comunicacional e inteligencia emocional.

### Aportes de la tesis

El presente trabajo es una propuesta de formación para un grupo de personas con discapacidad (PCD) emergente gracias a los avances de la tecnología - implantes cocleares y los audífonos de alta gama, sistemas de FM y demás dispositivos tecnológicos en constante desarrollo - que han permitido a muchas personas hipoacúsicas moderadas y severas poder seguir oyendo y aspirar a una mayor inclusión social, laboral y educativa.

### Líneas de I/D futuras

Una posible línea de investigación consiste en ampliar el número de personas que recorran el SETH, con el fin de obtener una perspectiva más amplia acerca de las posibles mejoras o ampliación de temáticas de interés.

En cuanto a los aspectos técnicos del SETH, se prevé una etapa de modificaciones conforme a los resultados obtenidos en las evaluaciones realizadas y a los avances tecnológicos futuros.



**Advisor**  
Mg. Alejandro González  
**Scientist Consultant**  
Dra. Alcira Vallejo  
**Thesis approval date**  
March 16, 2022

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

## **Strategies for the implementation of a job interview simulator for adults with hearing loss using hearing aids**

**Keywords:** Hearing impaired; Educational simulators; Technology applied in education

### **Motivation**

This thesis develops an online simulator prototype as an inclusive job interview training tool for hearing impaired adults.

#### **Specific Aims**

(confidence in communication- emotional intelligence)

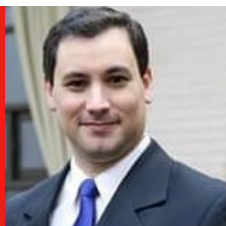
- Analyze the strong and weak points in the process of new knowledge acquisition by postlocutive hearing impaired adults.
- Study the concept of online simulator and its basis in the education and training context.
- Develop an online simulator prototype to train people how to face a job interview successfully.
- Evaluate the effectiveness of the skills which stimulate attention, memory, communication strategies and emotional intelligence.

### **Thesis contributions**

This work is a training proposal for a group of hearing disabled people emerging due to technological development. Cochlear implants, highly developed hearing aids, FM systems and other developing technological devices have given the opportunity to people with moderate and severe hearing loss to hear and aspire to more social, laboral and educational inclusion.

### **Future Research Lines**

A possible issue of development consists of increasing the number of SETH users with the purpose of collecting further information to improve the virtual environment and include in it possible topics of interest manifested by the target audience. Modifications are considered from the technical point of view of the SETH according to the results obtained from the evaluations given and the technological development.



**Directora**  
Dra. Cecilia Sanz

**Codirectora**  
Esp. Gladys Gorga

**Fecha de aprobación**  
6 de julio de 2022

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

## **Realidad Aumentada en Contextos Educativos y su Relación con el Rendimiento Académico Universitario**

**Palabras clave:** Realidad Aumentada; Rendimiento Académico; Innovación Educativa; Tecnología Informática Aplicada a Educación

### **Motivación**

La Realidad Aumentada (RA) viene demostrando beneficios en diversos escenarios educativos. Algunos de ellos se relacionan con una mejora en la comprensión de temas abstractos y en el desarrollo de habilidades. También en la literatura se mencionan mejoras sobre algunas variables tales como: motivación, interés, atención, satisfacción, interacción, sensación de control, entre otras, vinculadas al rendimiento académico, al utilizar estrategias pedagógicas con materiales con Realidad Aumentada.

El objetivo general de este trabajo es investigar acerca de la aplicación de Realidad Aumentada, como tecnología emergente, en diferentes contextos educativos y su vinculación con variables que influyen en el rendimiento académico. Particularmente, se analizan estudios acerca de la relación que existe entre RA y el rendimiento académico en el aula universitaria.

Como objetivos específicos se plantean:

- Investigar acerca de la utilización de RA en los diferentes contextos educativos.
- Indagar acerca de los aspectos relacionados con el rendimiento académico.
- Investigar acerca de la vinculación que existe entre la RA y diferentes variables que influyen en el rendimiento académico.
- Diseñar una experiencia que utilice tecnologías con RA para su aplicación en el ámbito de la enseñanza universitaria.
- Realizar un estudio de caso para visibilizar alcance, limitaciones y potencialidades de dicha tecnología.

### **Aportes de la tesis**

Este trabajo de tesis analiza la aplicación de la Realidad Aumentada en diferentes contextos educativos. Se consideran sus características, los elementos de hardware y software necesarios para su utilización, y su encuadre como herramienta, que mediante diferentes estrategias pedagógicas, la incorporan en distintos niveles educativos.

Como foco principal se estudian las vinculaciones entre la Realidad Aumentada y el rendimiento académico de los estudiantes considerando algunas variables que se asocian a éste. Para ello, se realizó una revisión sistemática de la literatura, primeramente del concepto de rendimiento y sus variables, y posteriormente de trabajos que integran RA e investigan su impacto en el rendimiento académico.

A partir de esta investigación se decidió considerar como variables de estudio la motivación, atención, interés, satisfacción, y calificación, todas ellas relacionadas al rendimiento académico.

Como aporte principal se diseñó un material educativo con RA que integra un “asistente virtual” para acompañar al estudiante con feedback sobre cómo realizar los ejercicios de una práctica dada como tarea. Además, se llevó a cabo una experiencia en el ámbito de la enseñanza universitaria con el material diseñado, que conformó el estudio de caso planteado en la propuesta de esta tesis. El objetivo del estudio de caso se centró en indagar la relación entre la utilización de material con RA y la mejora en el rendimiento académico, considerando las variables antes mencionadas.

Los resultados obtenidos a partir del estudio de caso, permiten confirmar algunos de los aportes encontrados a partir

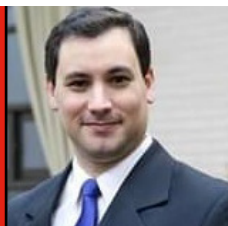
de la investigación teórica, y afirmar que la utilización de tecnologías de RA, bajo un encuadre pedagógico adecuado, tiene un impacto positivo tanto en los procesos de enseñanza y aprendizaje como en las variables objeto de estudio que inciden en la mejora del rendimiento académico de los estudiantes.

### **Líneas de I/D futuras**

Esta investigación obtiene resultados similares a los encontrados en la revisión sistemática de la literatura, sin embargo, la muestra no permite ser concluyente por lo que se continuará profundizando con nuevos estudios de casos en diferentes contextos. Al mismo tiempo, se abre la puerta para investigar sobre algunas variables de entrada que también pueden incidir en los resultados tales como género, edad y conocimientos previos.

Debido a una limitación encontrada durante el estudio de caso, referida al tipo de dispositivo con los que contaban los estudiantes, se propone continuar con la investigación de herramientas de autor que posibiliten la creación de materiales aumentados con fines educativos para diferentes plataformas.

Por último, se espera avanzar en el diseño y desarrollo de materiales educativos con tecnologías con Realidad Aumentada en otras cátedras en el ámbito universitario, con el uso de un “asistente virtual”, que acompañe a los estudiantes en sus tareas.



**Advisor**  
Dra. Cecilia Sanz

**Codirector**  
Esp. Gladys Gorga

**Thesis approval date**  
July 6, 2022

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

## **Augmented Reality in Educational Contexts and its Relationship with University Academic Performance**

**Keywords:** Augmented Reality; Academic performance; Educational Innovation; Computer Technology Applied to Education

### **Motivation**

Augmented Reality (AR) has been demonstrating benefits in various educational scenarios. Some of them are related to an improvement in the understanding of abstract topics and in the development of skills. The literature also mentions improvements on some variables such as: motivation, interest, attention, satisfaction, interaction, sense of control, among others, linked to academic performance, when using pedagogical strategies with Augmented Reality materials.

The general objective of this work is to investigate the application of Augmented Reality, as an emerging technology, in different educational contexts and its relationship with variables that influence academic performance. Particularly, studies about the relationship between AR and academic performance in the university classroom are analyzed.

As specific objectives we propose:

- To investigate about the use of AR in different educational contexts.
- To inquire about the aspects related to academic performance.
- To investigate the link between AR and different variables that influence academic performance.
- To design an experience that uses AR technologies for its application in the university teaching environment.
- Conduct a case study to visualize the scope, limitations and potential of this technology.

### **Thesis contributions**

This thesis work analyzes the application of Augmented Reality in different educational contexts. It considers its characteristics, the hardware and software elements necessary for its use, and its framing as a tool that, through different pedagogical strategies, incorporates it in different educational levels.

The main focus is to study the links between Augmented Reality and the academic performance of students considering some variables associated with it. For this purpose, a systematic review of the literature was carried out, first of the concept of performance and its variables, and then of works that integrate AR and investigate its impact on academic performance.

Based on this research, it was decided to consider motivation, attention, interest, satisfaction, and grade, all of them related to academic performance, as study variables.

As a main contribution, an educational material was designed with AR that integrates a “virtual assistant” to accompany the student with feedback on how to perform the exercises of a given practice as homework. In addition, an experience was carried out in the field of university teaching with the designed material, which formed the case study proposed in the proposal of this thesis. The objective of the case study focused on investigating the relationship between the use of AR material and the improvement in academic performance, considering the

variables mentioned above.

The results obtained from the case study confirm some of the contributions found from the theoretical research, and affirm that the use of AR technologies, under an adequate pedagogical framework, has a positive impact both on the teaching and learning processes and on the variables under study that have an impact on the improvement of students' academic performance.

### **Future Research Lines**

This research obtains results similar to those found in the systematic review of the literature; however, the sample does not allow us to be conclusive, so we will continue to deepen with new case studies in different contexts. At the same time, it opens the door to investigate some input variables that may also affect the results, such as gender, age and previous knowledge.

Due to a limitation found during the case study, referred to the type of device the students had, it is proposed to continue with the research of authoring tools that enable the creation of augmented materials for educational purposes for different platforms.

Finally, it is expected to advance in the design and development of educational materials with Augmented Reality technologies in other university courses, with the use of a "virtual assistant" to accompany students in their tasks.



Directora  
Dra. Roxana Giandini

Asesora Científica  
Dra. María de los Ángeles Bacigalupe

Fecha de aprobación  
2 de diciembre de 2022

SEDICI  
[Tesis Completa](#)

## Deserción estudiantil en la UTN-FRLP. Detección temprana mediante un sistema software de gestión tutorial

**Palabras clave:** Sistemas de Gestión Tutorial; Deserción Estudiantil; Deserción Universitaria; Tutorías; Indicadores de Deserción; Indicadores de Riesgo de Deserción

### Motivación

Las estadísticas de la Organización para la Cooperación y el Desarrollo Económicos (OCDE) indican que la educación superior argentina presenta una alta tasa de matriculación, semejante a la que se observa en varios países europeos. Sin embargo, esta elevada tasa de escolarización tiene su contracara con una alta tasa de deserción. Argentina es uno de los países con menor graduación en proporción al total de estudiantes universitarios tanto en Latinoamérica como a nivel mundial.

Como parte de las instituciones educativas, la UTN Facultad Regional La Plata (UTN-FRLP) no está exenta a esta problemática. Por este motivo, desde principios del año 2017 se encuentra trabajando en el Proyecto de Investigación y Desarrollo (PID) denominado “Estudio Sistemático de Deserción Estudiantil Universitaria” homologado por Rectorado UTN; el cual continúa su línea de trabajo a partir del año 2020 con nuevos proyectos vinculados a la generación de herramientas informáticas que asistan en el abordaje de la deserción estudiantil.

Dentro de esta línea de investigación surge la presente tesis que plantea como objetivo general el diseño y desarrollo de un sistema software de gestión tutorial. Sus funciones se basan en calcular automáticamente los factores de riesgo de deserción, presentar alertas tempranas y gestionar de manera eficiente el seguimiento de los estudiantes, facilitando el conjunto de tareas de gestión tutorial.

Para realizar el cálculo del riesgo de deserción utiliza una fórmula matemática fundamentada en los principales enfoques que conforman el marco teórico y el análisis particular del perfil de los estudiantes desertores de la institución, conformando tres pilares o ejes de análisis: Perfil Socioeconómico, Inasistencias y Rendimiento Académico.

### Aportes de la tesis

El sistema propuesto se desarrolla como una herramienta fundamental para el Programa de Acompañamiento de las Trayectorias Educativas (PATE) que tiene lugar en el ámbito de la Secretaría Académica UTN-FRLP. Este Programa surge de la necesidad de generar políticas de acompañamiento para los estudiantes en su recorrido por la vida universitaria, en el cual “Trayectorias Continuas” es uno de ejes centrales. Dentro de este espacio se trabaja sobre las experiencias que surgen alrededor del ámbito académico e institucional y que representan momentos relevantes como el acceso, el tránsito por la carrera, la permanencia, la interrupción temporal y el egreso con la titulación universitaria. En este contexto, el sistema propuesto brinda un aporte sustancial a las actividades desarrolladas dentro de este programa al constituir una herramienta informática que brinda soporte integral al programa PATE.

En consecuencia, se esperan como aportes del proyecto:

- (i) Promover las intervenciones de tutores facilitando la aplicación de estrategias particularizadas a las necesidades de cada estudiante.
- (ii) Constituir el primer estudio formal sobre deserción en la UTN-FRLP.
- (iii) Brindar un modelo de detección temprana de deserción estudiantil que contribuya a la mejora de los indicadores académicos en general.
- (iv) Visibilizar indicadores que sirvan de comparación con otras instituciones ofreciendo un aporte a la comunidad académica universitaria en general.
- (v) Promover el incremento en las tasas de promoción efectiva, lo cual redundará directamente en la reducción de la deserción universitaria. De esta manera, contribuye con las Políticas Públicas de transformación social en el área de la Educación Superior. Asimismo y en este sentido, espera brindar sustento a los



objetivos fundacionales de la Universidad Obrera Nacional que dio origen a la actual Universidad Tecnológica Nacional; la cual tenía como objetivo primario impartir formación integral a los profesionales de origen obrero, acentuando su papel en la construcción de la educación universitaria, y con él, asumiendo una participación activa en el desarrollo de la industria.

## Líneas de I/D futuras

### Nuevas funcionalidades:

Se prevé en versiones futuras del sistema incorporar nuevas funcionalidades, que se detallan a continuación:

(i) Los tutores podrán ajustar el semáforo asociado a un estudiante cuando el semáforo calculado por el sistema no refleje la condición particular del mismo.

(ii) Los estudiantes podrán solicitar cambio de tutor desde AlumnosWeb (Sistema Web de trámites estudiantiles de la institución).

(iii) Será privilegio de un usuario administrador poder modificar la fórmula del semáforo automático, es decir, actualizar los indicadores que conforman cada pilar y la ponderación asociada a cada uno de los mismos.

(iv) El sistema contemplará notificaciones en pantalla y vía email, de eventos y contactos semanales programados para estudiantes, tutores y coordinadores de tutores.

(v) La incorporación de una mensajería permitirá intercambiar mensajes asincrónicos entre tutores, coordinador de tutores y estudiantes agilizando la comunicación entre los distintos actores.

### Tablero estadístico:

Se prevé incorporar una nueva herramienta de visualización de datos que ofrezca a las autoridades académicas, información sintética y dinámica sobre los índices de deserción, rendimiento académico y datos socioeconómicos de los estudiantes y graduados.

### Transferencia a instituciones:

Se prevé iniciar contactos con instituciones educativas con el fin de aportar conocimientos y experiencias en proyectos vinculados con tecnología informática aplicada al abordaje de la deserción estudiantil.

### Implementación en otras instituciones:

Debido a las características técnicas con las cuales se ha desarrollado el sistema es factible su aplicación en nuevos ámbitos de aplicación, resultando su aplicación directa en otras facultades regionales de la UTN, como así también en otras instituciones de educación superior o media.

### Incorporación de indicadores obtenidos de entornos virtuales de enseñanza-aprendizaje (EVEAS):

Motivado por las nuevas modalidades educativas por la pandemia relacionada con el covid-19 y teniendo en cuenta que el sistema comenzó su desarrollo en un momento en el que las clases presenciales eran parte de la única modalidad

brindada por la institución, se plantea como trabajo futuro la incorporación de indicadores del nuevo concepto de presencialidad, regulación del aprendizaje y rendimiento académico en entornos virtuales de enseñanza-aprendizaje (EVEAS).

Se espera, de esta manera, que el sistema amplíe su modelo de indicadores de riesgo de deserción estudiantil incorporando la posibilidad de tomar datos relacionados tanto con modelos educativos presenciales, como a distancia y semipresenciales, incorporando además la evaluación de indicadores asociados a la regulación del aprendizaje. Siguiendo este objetivo, en el año 2022 da inicio un proyecto denominado "Sistema informático de predicción de deserción estudiantil universitaria basado en un modelo de indicadores de regulación del aprendizaje, en entornos educativos mediados por TIC" homologado por Rectorado UTN con código SIIPBLP0008384, el cual persigue los objetivos planteados complementando el trabajo precedente propuesto en esta tesis.



**Advisor**

Dra. Roxana Giandini

**Scientist Consultant**

Dra. María de los Ángeles Bacigalupe

**Thesis approval date**

December 2, 2022

**SEDICI**

[Full Text](#)

**Student Desertion from the UTN-FRLP.  
Early detection using a tutorial  
management software**

**Keywords:** Tutorial systems; student desertion; undergraduate desertion; tutoring; desertion indicators; desertion risk indicators

## Motivation

The statistics from the Organization for Economic Co-operation and Development (OCDE, Spanish acronym) indicate that the Argentinian high school education system has a high enrolment rate, as it happens in several European countries. However, this high rate of students entails a high rate of desertion.

Argentina is one of the countries with less graduates in relation to the total number of undergraduates in Latin America and worldwide.

As a member of the education institutions, the UTN -La Plata Region (UTN-FRLP)- also has the same problem. For this reason, since the beginning of 2017, the thesis writer is part of the I&D Project (PID) called "Systematic study of undergraduate desertion" officially authorized by the UTN President's Office. This project continued in 2020 with new projects in connection to the creation of computer tools that help student desertion. Following this line of investigation, the present thesis work arises and aims to design and develop a tutorial software. The idea is that the software can automatically estimate the desertion risk indicators, give early signs and efficiently follow the undergraduates.

In order to calculate the desertion risk, it uses a mathematical formula based on the main approaches that form the theoretical frame and the specific analysis of the leaving students' profile. There are three central themes of the analysis: socio-economic profile, absences and academic performance.

## Thesis contributions

The system suggested is developed as an essential tool for the Programa de Acompañamiento de las Trayectorias Educativas (PATE, a programme that accompanies the students during their course of studies) which is part of

the Academic Secretary of the UTN-FRLP. This Programme is created to accompany the students during their university life, where "Continue trajectories" is one of the central themes. We work with what happens at university, important moments, such as, enrolment, the performance of the course of study, the stay, the temporal interruption, and the graduation.

In this context, the proposed system provides a substantial contribution to the activities comprised in this programme, as it is a computing tool that gives integral support to the PATE Programme.

So, the expected project contributions are the following:

- (i) Promote tutor's participation to ease the application of customized strategies for each student.
- (ii) Constitute the first formal study about desertion in UTN-FRLP.
- (iii) Improve the academic indicators, in general, by means of an early detection model of student desertion.
- (iv) Show indicators that help to establish a comparison with other institutions and contribute with the undergraduates in general.
- (v) Increase effective promotion rates so undergraduate desertion decreases. In this way, it contributes with the Public Policies of a social transformation at university. Besides, it wants to give support to the founding goals of the Universidad Obrera Nacional which has given origin to the current National Technological University; which had as a main goal to educate professionals who were coming from the working class, to put emphasis on the university education construction; and also, to be part of the industrial development.

## Future Research Lines

### New functionalities:

In the future, the idea is to incorporate the following functionalities:

- (i) Tutors can adjust the traffic light of a student if it does not reflect the student's condition.
- (ii) Students can ask for a change of tutor from AlumnosWeb (a web system created for undergraduates).
- (iii) The administrator can modify the formula of the automatic traffic light, i.e., update the indicators of each pillar and the information considered in each of them.
- (iv) The system will send notifications via e-mail of all the events and scheduled meetings to students, tutors and tutors' coordinators.
- (v) The e-mail incorporation enables asynchronous messages to communicate among tutors, their coordinators and students.

### Statistical board:

The system will include a new tool to visualize data to offer the academic authorities brief and dynamic information about desertion rates, academic performance and socio-economical profiles of students and graduates.

Moves to other institutions:

The idea is to establish relationships with different education institutions to share knowledge and experience in projects connected to computing technology applied to student desertion.

### Implementation in other institutions:

Due to the technical properties the systems has, it can be easily used in other regional schools of the UTN or in different institutions of middle or high education.

### Incorporation of indicators taken from virtual educating environments (EVEAS, Spanish acronym):

As a consequence of Covid, new methods of teaching have been created because face-to-face classes were the only ones at that time. So, in the future, it is necessary to incorporate indicators of this new concept of class-based, and regulate academic learning and performance in virtual learning environments (EVEAS).

We expect the system can enlarge its desertion risk indicator model and connect the information from the class-based, face-to-face and blended learning and evaluate the indicators of learning regulation.

Following this line, the year 2022 gave origin to a project called "Undergraduate desertion predictive system based on indicators of learning regulation in environments working with TICs" officially authorized by the UTN President's Office under number SIIPBLP0008384, which tries to accomplish the aims proposed together with the preceding thesis work.



**Directores**

Dr. José Luís Hernández  
Lic. Francisco Javier Díaz

**Fecha de aprobación**

13 de diciembre de 2022

**SEDICI**

[Tesis Completa](#)

## Optimización de Tráfico en Redes Multiservicios Aplicando Técnicas Heurísticas

**Palabras clave:** MPLS; Optimización; Algoritmos bio-inspirados; Optimización por colonia de hormigas; Optimización por enjambres de partículas; Optimización por algoritmo de murciélagos.

### Motivación

Los nuevos requerimientos de tráfico de las redes modernas donde el histórico control basado en el mejor esfuerzo no resulta eficiente para evitar la congestión, y la implicancia de adecuar los modelos de tráfico mediante la aplicación de técnicas de optimización que ayuden a la consecución del objetivo de utilizar eficientemente los recursos de la red, motivaron el desarrollo de este trabajo.

MPLS fue diseñado para atender las necesidades de las redes modernas y han recibido gran interés en los últimos años debido a las posibilidades, flexibilidad y eficiencia que brinda en ingeniería de tráfico.

En esta tesis se propone, mediante técnicas no determinísticas, resolver el problema de asignación de recursos frente a situaciones de múltiples requerimientos en una red MPLS multicaminos, donde el objetivo es minimizar el costo de ruteo, cumpliendo con restricciones de capacidad de los enlaces y cumpliendo con demandas específicas para una planificación de la red fuera de línea.

El objetivo general de esta tesis es resolver el problema de optimización de asignación de recursos en redes multiservicios mediante la aplicación de técnicas heurísticas y meta-heurísticas puras o combinadas.

Los objetivos particulares son los siguientes:

- Contribuir a completar una taxonomía sobre la base del análisis de diferentes estrategias heurísticas puestas a resolver problemas combinatorios en el ámbito de las redes de comunicaciones.
- Diseñar e implementar una batería de algoritmos que permitan resolver casos diferentes que respondan a modelado similar.
- Estudiar sobre la combinación de heurísticas y determinar un criterio de selección para cada tipo de problema considerado.

### Aportes de la tesis

Esta tesis se ha centrado, concibiendo a MPLS como una técnica orientada a la conexión, en el problema de ingeniería de tráfico fuera de línea. Luego de entender el funcionamiento, características y las ventajas indiscutibles de MPLS se realizó un modelado matemático de la red que permitió establecer mecanismos para abordar el problema de la asignación de recursos en situaciones de múltiples requerimientos en una red multicaminos.

Determinada la complejidad del problema y descartando la utilidad de métodos deterministas que pudieran abordarlo, se sintetizó una taxonomía de estrategias heurísticas y se diseñaron e implementaron cinco algoritmos bioinspirados en enjambres diferentes para resolver el problema de la asignación óptima de flujos en redes MPLS.

En general, las implementaciones de metaheurísticas basadas en el comportamiento de seres vivos diseñadas para resolver un problema concreto, como en el caso de la presente tesis, necesitan la construcción de estrategias basadas en las metáforas del comportamiento de los organismos vivos, pero necesariamente adaptadas al problema que se resuelve. La hibridación de varios métodos y la inclusión de conocimiento específico del problema se deben combinar de manera tal que coadyuve a la consecución de los requerimientos que la ingeniería exige: obtener buenas soluciones en tiempos aceptables.

De los cinco algoritmos desarrollados, uno está inspirado en bandadas de pájaros (PSO), dos algoritmos están inspirados en el comportamiento de colonias de hormigas (ACO 1 y ACO 2) uno inspirado en el comportamiento de los murciélagos y el último una hibridación entre el comportamiento de los murciélagos y las hormigas. La diferencia entre ellos recae

en la forma en que se desplazan por el espacio de búsqueda, lo que resalta las ventajas de unos u otros dependiendo de las situaciones o instancias a resolver. La combinación entre ellos o su hibridación, resulta en una mejora significativa de los resultados obtenidos.

Experimentalmente se ratificó el funcionamiento de los algoritmos tanto para instancias pequeñas del problema, con excelente desempeño, como para instancias medianas, grandes y muy grandes, en las cuales se pusieron de manifiesto las ventajas de cada implementación. Mientras algunos entregaron soluciones en tiempos menores, otros tuvieron la ventaja de realizar un mejor balance de las cargas entre los caminos posibles.

Se considera un aporte importante de esta tesis al hecho de que la mayoría de las estrategias heurísticas descritas podrán adaptarse para la resolución de otros problemas de optimización de asignación de recursos con relativa facilidad.

Por otra parte, el modelo matemático de la red puede ser reformulado en forma sencilla no solo utilizando distintas métricas que evalúen su costo, sino focalizando diferentes aspectos a optimizar. Podrían introducirse aspectos tales como funciones de equilibrio de carga específicos, mecanismos de control de congestión, políticas distintas para minimizar el costo de ruteo, distintas políticas de enrutamiento, etc.

Se ha formulado un problema de gran alcance, se han resuelto instancias que han permitido testear un conjunto de estrategias utilizadas en forma pura e híbridas. Se han logrado encontrar en todos los casos buenas soluciones en tiempos aceptables constituyendo una propuesta válida para el campo de la ingeniería.

### Líneas de I/D futuras

Se propone como investigación y desarrollos futuros:

- La incorporación de otras estrategias para ensayar nuevas hibridaciones, particularmente las relacionadas con la incorporación de búsqueda local como mecanismo de explotación de regiones promisorias del espacio de búsqueda.
- La aplicación de los algoritmos propuestos a la resolución de otro tipo de problemas de asignación de recursos, considerando diferentes funciones objetivo, que incluyan métricas más complejas y nuevos objetivos, incluso contrapuestos a los planteados en este trabajo dando lugar a un análisis conducente a la resolución de problemas de optimización combinatoria multiobjetivo.
- Abordar el diseño de interfaces que permitan trasladar las soluciones obtenidas a los dispositivos de la red de manera sencilla y/o automática, por ejemplo, generando los archivos de configuración de los equipos de la red en producción.
- Aplicación de las técnicas descritas a SDN. Si bien los sistemas SDN hacen al cliente menos dependiente de un proveedor MPLS, estas tecnologías permiten al proveedor de servicios aumentar y extender las implementaciones MPLS existentes.

Esta combinación de los mecanismos MPLS y la arquitectura SDN permiten la automatización de la red y sus operaciones a través de un control centralizado lo que facilita su optimización y planificación. Esto plantea un desafío interesante para continuar con la investigación, desarrollo e innovación de técnicas que permitan optimizar el funcionamiento de las arquitecturas de red del futuro, para garantizar métodos rentables y ágiles que aseguren la plena simbiosis con los avances permanentes y sostenidos en TIC a través de la red.



**Advisors**

Dr. José Luíz Hernández  
Lic. Francisco Javier Díaz

**Thesis approval date**

December 13, 2022

SEDICI

[Full Text](#)

## Traffic optimization in multiservice networks using heuristic techniques

**Keywords:** MPLS; Optimization; Bio-inspired algorithms; Ant colony optimization; Particle swarm optimization; Bat algorithm optimization

### Motivation

The new modern networks traffic requirements where the historical best-effort control is not efficient to avoid congestion, and the implication of adapting traffic models through the optimization techniques application that help to achieve the goal of efficient use of network resources, motivated the development of this work.

MPLS was designed to meet the needs of modern networks and has received great interest in recent years due to the possibilities, flexibility and efficiency it offers in traffic engineering.

In this thesis we propose, using nondeterministic techniques, to solve the resource allocation problem in a multipath MPLS network, where the objective is to minimize the routing cost, meeting link capacity constraints and complying with specific demands for offline network planning.

The general objective of this thesis is to solve the resource allocation optimization problem in multiservice networks by applying pure or combined heuristic and meta-heuristic techniques.

The particular objectives are the following:

- To contribute to complete a taxonomy on the basis of the analysis of different heuristic strategies set to solve combinatorial problems in the field of communication networks.
- To design and implement a battery of algorithms to solve different cases responding to similar modeling.
- Study on the combination of heuristics and determine a selection criterion for each type of problem considered.

### Thesis contributions

This thesis has focused, conceiving MPLS as a connection-oriented technique, on the problem of off-line traffic engineering. After understanding the operation, characteristics and undisputed MPLS advantages, a network mathematical modeling was performed to establish mechanisms to address the problem of resource allocation in multiple requirements situations in a multipath network.

Determining the complexity of the problem and ruling out the usefulness of deterministic methods that could address it, a taxonomy of heuristic strategies was synthesized and five different swarm bio-inspired algorithms were designed and implemented to solve the optimal flow allocation problem in MPLS networks.

In general, metaheuristics implementations based on the living beings behavior designed to solve a particular problem, as in the case of the present thesis, require the construction of strategies based on the living organisms behavior metaphors but necessarily adapted to the problem being solved. The several methods hybridization and the specific knowledge inclusion of the problem must be combined in such a way that it contributes to the achievement of the requirements that engineering demands: to obtain good solutions in acceptable times.

Five algorithms were developed, one is inspired by bird flocks (PSO), two algorithms are inspired by the ant colonies behavior (ACO 1 and ACO 2), one is inspired by the bats behavior and the last one is a hybridization between the bats and ants behavior. The difference between them lies in the way they move through the search space, which

highlights the advantages of one or the other depending on the situations or instances to be solved. The combination between them or their hybridization results in a significant improvement of the results obtained.

Experimentally, the operation of the algorithms was ratified for small instances of the problem, with excellent performance, as well as for medium, large and very large instances, in which the each implementation advantages became evident. While some delivered solutions in shorter times, others had the advantage of performing better load balancing among the possible paths.

It is considered an important contribution of this thesis that most of the heuristic strategies described can be easily adapted to solve other resource allocation optimization problems.

On the other hand, the mathematical model of the network can be reformulated in a simple way not only by using different metrics to evaluate its cost, but also by focusing on different aspects to be optimized. Aspects such as specific load balancing functions, congestion control mechanisms, different minimize routing cost policies, different routing policies, etc., could be introduced.

A wide-ranging problem has been formulated, instances have been solved and a set of strategies used in pure and hybrid form have been tested. In all cases, good solutions have been found in acceptable times, constituting a valid proposal for the engineering field.

### Future Research Lines

It is proposed as future research and development:

- The other strategies incorporation to test new hybridizations, particularly those related to the incorporation of local search as a mechanism to exploit promising regions of the search space.
- The application of the proposed algorithms to the resolution of other types of resource allocation problems, considering different objective functions, including more complex metrics and new objectives, even opposed to those proposed in this work, leading to an analysis conducive to the resolution of multi-objective combinatorial optimization problems.
- Addressing the design of interfaces that allow to transfer the obtained solutions to the network devices in a simple and/or automatic way, for example, generating the configuration files of the network equipment.
- Application of the described techniques to SDN. While SDN systems make the customer less dependent on an MPLS provider, these technologies allow the service provider to augment and extend existing MPLS deployments. This combination of MPLS mechanisms and SDN architecture

enables automation of the network and its operations through centralized control for easy optimization and planning. This poses an interesting challenge for further research, development and innovation techniques to optimize the future network architectures operation, to ensure cost-effective and agile methods to ensure full symbiosis with the permanent and sustained advances in ICT across the network.





# 03

ESPECIALIZACIONES

INGENIERÍA DE SOFTWARE

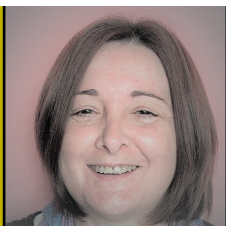
INTELIGENCIA DE DATOS  
ORIENTADA A BIG DATA

CÓMPUTO DE ALTAS PRESTACIONES  
Y TECNOLOGÍA GRID

COMPUTACIÓN GRÁFICA,  
IMÁGENES Y VISIÓN POR  
COMPUTADORA

TECNOLOGÍA INFORMÁTICA  
APLICADA EN EDUCACIÓN

REDES Y SEGURIDAD



**Director**  
Dr. Leandro Antonelli

**Codirector**  
Mg. Pablo Thomas

**Fecha de aprobación**  
1 de septiembre de 2022

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

## **Un análisis de enfoques de Validación de Requerimientos**

**Palabras clave:** Ingeniería de requerimientos; Validación de requerimientos; Enfoques de validación de requerimientos

### **Motivación**

“La parte más difícil de crear un sistema de software es decidir exactamente qué construir. Ninguna otra parte del trabajo conceptual es tan difícil como establecer los requerimientos técnicos detallados...Ninguna otra parte del trabajo paraliza tanto el sistema resultante si se hace mal. Ninguna otra parte están difícil de rectificar posteriormente”.

Uno de los factores más importantes en el desarrollo de un proyecto de software es Validación de Requerimientos. Los requerimientos erróneos, si no se detectan a tiempo, pueden causar graves problemas, como costos adicionales, incumplimiento de los objetivos esperados y retrasos en las fechas de entrega. Por estas razones, se debe dedicar un gran esfuerzo a la Ingeniería de Requerimientos para garantizar que los requerimientos del proyecto sean acordes al problema a resolver. Esto permite vislumbrar la importancia que tiene para la calidad del software implementar en cada proyecto de desarrollo de software una correcta Ingeniería de Requerimientos. La detección de un error en la etapa posterior del desarrollo de software suele llevar más tiempo y es más costoso que si el error se detecta en la etapa inicial. Por ejemplo, Boehm y Basili indican que encontrar y corregir defectos después del lanzamiento de un producto es a menudo 100 veces más costoso que corregirlos durante la fase de requerimientos y modelado. Además, una encuesta realizada por Hofmann y Lehner muestra que los proyectos de software exitosos han invertido alrededor del 28% del esfuerzo en la fase de inicial de requerimientos.

Una actividad importante y necesaria en el proceso de Ingeniería de Requerimientos que puede evitar que los errores se propaguen a una fase de desarrollo posterior es la Validación de Requerimientos. La Validación de Requerimientos es una

tarea difícil porque, por naturaleza, una especificación de requerimientos puede ser interpretada de diferentes maneras y en diferentes contextos por las partes interesadas. Hasta la fecha, diferentes investigadores han propuesto y desarrollado numerosas técnicas de para mejorar los procesos de Validación de Requerimientos y su eficacia, como revisiones, inspecciones, creación de prototipos y métodos formales. Cada una de estas técnicas posee fortalezas y debilidades inherentes sobre otras, por lo que la elección de las técnicas de validación de requerimientos debe realizarse considerando el tipo de sistemas que se está desarrollando y los procesos de desarrollo de software de las organizaciones.

En el marco de la Ingeniería de Requerimientos la Validación de Requerimientos es una tarea fundamental en cualquier proyecto de Ingeniería de Software y debe ser un proceso continuo en el ciclo de vida del desarrollo del sistema. El principal objetivo de la Validación de Requerimientos es confirmar que los requerimientos especificados sean representaciones de las necesidades y expectativas de los usuarios y que además sean completos, correctos y consistentes, entre otras características.

Según Kotonya la Validación de Requerimientos se refiere a verificar la coherencia, integridad y corrección del documento de requerimientos, también se establece que los requerimientos deben ser: validos, comprensibles, consistentes, trazables, íntegros, reales y verificables.

Trabajar en la Validación de Requerimientos se está convirtiendo en un desafío para los equipos de desarrollo de software, clientes y usuarios. Existen diferentes causas que imponen problemas de comunicación, control, intercambio de conocimientos, confianza y retrasos en el desarrollo del software.

Este trabajo se enfoca en las principales tendencias de la Validación de Requerimientos de software a través de una evaluación de distintos enfoques con el objetivo de establecer un marco conceptual y referencial del tema.

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

El presente trabajo tiene como objetivo relevar, analizar y comparar los distintos enfoques existentes en la literatura para la validación de requerimientos de software, es decir, establecer un marco referencial y conceptual del tema. Como resultado de este trabajo de especialización se espera brindar una caracterización de los enfoques existentes, de manera tal que los profesionales de la disciplina cuenten con una herramienta para decidir que enfoque utilizar.

### **Líneas de I/D futuras**

Se propone como trabajo futuro desarrollar un enfoque que puede ayudar en el proceso de validación de requerimientos. El enfoque a desarrollar, en primer lugar, debe describir un marco de referencia para el desarrollo de la estructura y en segundo lugar establecer un modelo para el proceso de validación de requerimientos a nivel de caja blanca en la estructura interna del diseño de los componentes del desarrollo del sistema. Esto permite garantizar la calidad del producto o sistema, enfocándose en las diferentes etapas del ciclo de vida del software.



**Advisor**  
Dr. Leandro Antonelli

**Codirector**  
Mg. Pablo Thomas

**Thesis approval date**  
Septemeber 1st, 2022

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

## **An Analysis of Requirements Validation Approaches**

**Keywords:** Requirements engineering; Requirements validation; Requirements validation approaches.

### **Motivation**

“The hardest part of creating a software system is deciding exactly what to build. No other part of the conceptual work is as difficult as establishing the detailed technical requirements...No other part of the work cripples the resulting system as much if it is done wrong. No other part is as difficult to rectify later.”

One of the most important factors in the development of a software project is Requirements Validation. Wrong requirements, if not detected in time, can cause serious problems, such as additional costs, failure to meet expected objectives and delays in delivery dates. For these reasons, a great effort must be devoted to Requirements Engineering to ensure that the project requirements are in line with the problem to be solved. This gives a glimpse of how important it is for software quality to implement a correct Requirements Engineering in every software development project.

Detecting a bug at the later stage of software development is often more time-consuming and costly than if the bug is detected at the initial stage. For example, Boehm and Basili indicate that finding and correcting defects after a product release is often 100 times more costly than correcting them during the requirements and modeling phase. In addition, a survey by Hofmann and Lehner shows that successful software projects have invested about 28% of the effort in the initial requirements phase.

An important and necessary activity in the Requirements Engineering process that can prevent errors from propagating to a later development phase is Requirements Validation. Requirements Validation is a difficult task because, by nature, a requirements specification can be interpreted in different ways and in different contexts by

interested parties. To date, numerous techniques have been proposed and developed by different researchers to improve Requirements Validation processes and their effectiveness, such as reviews, inspections, prototyping and formal methods. Each of these techniques has inherent strengths and weaknesses over others, so the choice of requirements validation techniques must be made considering the type of systems being developed and the software development processes of the organizations.

In the framework of Requirements Engineering, Requirements Validation is a fundamental task in any Software Engineering project and should be a continuous process in the system development life cycle. The main objective of Requirements Validation is to confirm that the specified requirements are representations of the users' needs and expectations and that they are complete, correct and consistent, among other characteristics.

Según Kotonya la Validación de Requerimientos se refiere a verificar la coherencia, integridad y corrección del documento de requerimientos, también se establece que los requerimientos deben ser: validos, comprensibles, consistentes, trazables, íntegros, reales y verificables.

Working on Requirements Validation is becoming a challenge for software development teams, customers and users. There are different causes that impose problems of communication, control, knowledge sharing, trust and delays in software development.

This paper focuses on the main trends in software Requirements Validation through an evaluation of different approaches with the objective of establishing a conceptual and referential framework of the subject.

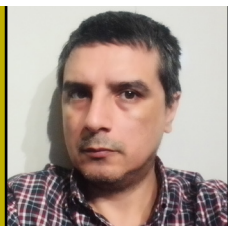
### **Thesis Final Work contributions**

The purpose of this work is to survey, analyze and compare the different approaches existing in the literature for the validation of software requirements, that is, to establish a referential and conceptual framework of the subject. As a result of this work of specialization, it is expected to provide a characterization of the existing approaches, so that the professionals of the discipline have a tool to decide which approach to use.

### **Future Research Lines**

It is proposed as future work to develop an approach that can assist in the requirements validation process. The approach to be developed, firstly, should describe a framework for the development of the structure and secondly establish a model for the requirements validation process at the white box level in the internal structure of the design of the components of the system development. This allows to guarantee the quality of the product or system, focusing on the different stages of the software life cycle.

## ESPECIALIZACIÓN INGENIERÍA DE SOFTWARE



**Esp. Malqui Fernandez Vidal**

**e-mail**

malkyvidal@gmail.com

español

**Director**  
Dr. Federico Balaguer

**Fecha de aprobación**  
18 de octubre de 2022

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

# La Programación Reactiva: Un nuevo enfoque para trabajar con código asíncrono en la programación web

**Palabras clave:** Programación Reactiva; Patrones de Diseño; Javascript; Programación Web; Manejo de Eventos

## Motivación

En la actualidad Javascript es uno de los lenguajes más utilizados en distintas aplicaciones tanto en browsers, computadoras de escritorio o dispositivos móviles. El manejo de eventos es una tarea clave en este tipo de soluciones. Pero el manejo imperativo de eventos en Javascript conduce a problemas de código con alto acoplamiento y baja reusabilidad. La programación reactiva propone abstracciones como el Observable -basados en conocidos patrones de diseño como Observer e Iterator- que permiten solucionar dichos problemas. El objetivo del presente trabajo es presentar un conjunto de problemas que se presentan en el manejo imperativo de eventos de UI en el lenguaje Javascript identificados a partir del análisis bibliográfico. Se compara el enfoque imperativo y procedural con su contrapartida en la Programación reactiva y se resaltan las soluciones reactivas en términos de la calidad del código resultante.

## Aportes del Trabajo Final Integrador

Se presentan tres problemáticas en la programación de eventos inherentes al enfoque imperativo presentes en el lenguaje Javascript: Variables condicionales, Backpressure y Wait for all. Se hace un análisis de abstracciones de la programación reactiva, sus vínculos con conocidos patrones de diseño y cómo dichas abstracciones permiten solucionar los problemas mencionados.

El enfoque reactivo es motorizado principalmente desde la industria, el presente trabajo brinda un aporte para entenderla desde una perspectiva más académica.

## Líneas de I/D futuras

### Refactorización a Programación Reactiva

Una línea de trabajo futuro consiste en la creación de herramientas de análisis estático que soporten refactoring de código imperativo que maneja eventos en JavaScript a su contraparte Reactiva. Estas herramientas permitirían la transformación de callbacks, estructuras condicionales y variables globales utilizados para manejar eventos de formulario o eventos Ajax en Observables y sus operadores.

### Análisis de Deuda Técnica:

Una de las dimensiones en que la deuda técnica puede manifestarse es mediante código repetitivo, duplicado o complejo. Esto es justamente el caso cuando se utiliza el enfoque imperativo para coordinar eventos en JavaScript.

Usando la información del código fuente (scripts) se pueden desarrollar herramientas de identificación automática de este tipo de defectos con foco en la identificación de estructuras condicionales y variables globales utilizados con la finalidad de coordinar eventos.





english

**Advisor**  
Dr. Federico Balaguer  
**Thesis approval date**  
December 18, 2022  
**SEDICI**  
[Full Text](#)

## **Reactive Programming: A new approach to working with asynchronous code in web programming**

**Keywords:** Reactive Programming; Design Patterns; Javascript; Web Programming; Event Management.

### **Motivation**

Javascript is currently one of the most used languages in different applications in browsers, desktop computers or mobile devices. Event handling is a key task in this type of solution. But imperative event handling in Javascript leads to problems with highly coupled and poorly reusable code. Reactive programming proposes abstractions such as the Observable -based on well-known design patterns such as Observer and Iterator- that allow these problems to be solved. The objective of this work is to present a set of problems that arise in the imperative handling of UI events in the Javascript language identified from the bibliographic analysis. The imperative and procedural approach is compared with its counterpart in Reactive Programming and reactive solutions are highlighted in terms of the quality of the resulting code.

### **Thesis Final Work contributions**

Three problems are presented in the programming of events inherent to the imperative approach present in the Javascript language: Conditional variables, Backpressure and Wait for all. An analysis of reactive programming abstractions is made, its links with well-known design patterns and how these abstractions allow to solve the mentioned problems. The reactive approach is driven mainly from the industry, the present work provides a contribution to understand it from a more academic perspective

### **Future Research Lines**

#### **Refactoring to Reactive Programming:**

A future line of work consists of creating static analysis tools that support refactoring from imperative code that handles events in JavaScript to its Reactive counterpart. These tools would allow the transformation of callbacks, conditional structures and global variables used to handle form events or Ajax events into Observables and their operators.

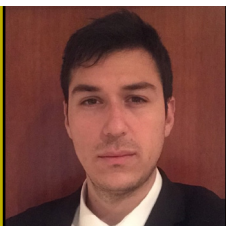
#### **Technical Debt Analysis:**

One of the dimensions in which technical debt can manifest itself is through boilerplate, duplicate, or complex code. This is precisely the case when using the imperative approach to coordinating events in JavaScript.

Using the information of the source code (scripts) it is possible to develop tools for automatic identification of this type of defects.

with focus on the identification of conditional structures and global variables used in order to coordinate events.

## ESPECIALIZACIÓN INTELIGENCIA DE DATOS ORIENTADA A BIG DATA



**Esp. Jonatan Emanuel Martinelli**

**e-mail**

jonatanmartinelli.e@gmail.com

español

**Advisor**

Dr. Waldo Hasperué

**Fecha de aprobación**

19 de septiembre de 2022

**SEDICI**

[Tesis Completa](#)

## **Clasificación de datos desbalanceados. Su aplicación en la predicción de bajas de beneficiarios de un servicio de salud privado**

**Palabras Clave:** Desbalance de clases ; Métodos de balanceo de clases ; Clasificación ; Machine Learning

### **Motivación**

Actualmente, los desarrollos más recientes de Inteligencia Artificial han resaltado la importancia de los sistemas de soporte de decisión en actividades empresariales, sugiriendo modelos desarrollados para asistir al tomador de resolución de problemas. Por lo tanto, el uso de herramientas de aprendizaje automático se ha incrementado en los últimos años, especialmente en aplicaciones del mundo real. Sin embargo, es claro que los datos reales no son perfectos, ya que pueden contener inconsistencias y valores faltantes, por lo que, dichos algoritmos deben ser sólidos para enfrentar las imperfecciones de los datos, y ser capaces de extraer patrones realmente fuertes.

El principal problema que presentan actualmente los algoritmos de aprendizaje automático, cuando son utilizados en tareas de clasificación con conjuntos de datos no balanceados, es el bajo desempeño alcanzado para clasificar correctamente ejemplos de la clase minoritaria (la menos representada). La mayoría de los trabajos previos que han tomado en cuenta el problema de clases no balanceadas en aplicaciones del mundo real, se han enfocado en evaluar el desempeño de los clasificadores, exclusivamente en términos de su capacidad para minimizar los errores de clasificación.

En este contexto y considerando el auge que ha tomado en la actualidad la ciencia de datos y las diversas tecnologías de información, el presente estudio, pretende determinar lo importante que resulta contar con un conjunto de datos equilibrado para la explotación de información la cual es útil para dar soporte a la toma de decisiones. Por tal motivo, se desarrollan modelos basados en diversas técnicas de nivelación de clases que permiten obtener conocimiento a partir de los datos generados por su sistema transaccional. De esta forma, el objetivo general de esta investigación es realizar un estudio, a partir de un conjunto de datos desbalanceado de una organización de cobertura médica, de diferentes estrategias de balanceo de clases para medir el poder pronóstico de diversos modelos obtenidos a través de dos técnicas de clasificación: Redes neuronales y Árboles de decisión.

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

Desde el punto de vista académico la investigación enriquece el conocimiento del alumno y de los lectores acerca de las variadas técnicas de balanceo de clases, así como también contar con modelos implementados bajo el proceso KDD y que sirvan de soporte a la gestión de una organización.

Desde el punto de vista teórico se considerará para la elaboración del modelo predictivo el proceso KDD para obtener un alto grado de eficacia y se investigarán las diferentes técnicas de balanceo de clases para mejorar la calidad de los datos.

De esta manera la organización espera planear las nuevas definiciones sobre el tratamiento de beneficiarios activos con nuevos enfoques, nuevas técnicas de retención y permitir a la alta gerencia tener información fehaciente, concreta y veraz, como soporte para tomar las decisiones más correctas. Por lo tanto, estos modelos tienen una influencia directa en la competitividad y en el proceso de planificación.

### **Líneas de I/D futuras**

Existen diversos algoritmos que en esta investigación no se han estudiado pero que pueden ser de alto interés para la cuestión del desbalance de clases, por ejemplo, el modelo logístico, la profundización de Remuestreo híbridos o el intervalo de confianza de un modelo. Algunos de ellos constaban de desarrollos metodológicos más complejos y con un tiempo de procesado superior a los que se han visto, y por este tipo de razones, no han sido considerados en la elaboración de este trabajo final integrador.

Por otra parte, no existe investigación alguna en este campo en otra área de la compañía, por lo que sería relevante para la gerencia comenzar a realizar algunas investigaciones que aclaren la función particular que desempeñan estas técnicas de Data Mining en las empresas. Como también es considerable la incentivación de futuras investigaciones dentro de esta área para obtener mejores resultados.



english

**Advisor**  
Dr. Hasperué Waldo  
**Thesis approval date**  
September 19, 2022  
**SEDICI**  
[Full Text](#)

**Classification of imbalanced datasets.  
Their application to predict the  
number of beneficiaries leaving  
a private health insurance entity**

**Keywords:** Class imbalance; How to deal with class imbalance; Classification; Machine learning

## Motivation

Nowadays, the latest developments of AI highlight the importance of decision support systems for business issues to help the problem solver. Thus, the implementation of machine learning for real-world applications has increased during these last years. However, it is clear that real data are not perfect; they can be inconsistent or incomplete so, those algorithms must be solid to tackle data imperfections and extract real strong patterns.

When these machine learning algorithms are used with imbalanced datasets to classify tasks, their performance to correctly classify examples from the minority class is really poor. Most of the previous works that have taken into account the problem of imbalanced classes in real-world applications have put emphasis on the performance of classifiers, in particular if they were capable of minimizing classification errors.

In this context, if we consider the boom of data science and of several information technologies, the aim of this present work is to demonstrate the great importance a balanced dataset to exploit information which is useful to give support to the decision-making processes. Following this line, we develop models based on different class balancing techniques that provide knowledge using the data obtained by the transactional system. Furthermore, the general goal of this research is to study -from an imbalanced dataset of a health insurance entity- the different class balancing strategies in order to measure the predictions of several models that work with two different classification techniques: neuronal networks and decision trees.

## Thesis Final Work contributions

From an academic point of view, research enriches the student's and readers' knowledge of the varied class balancing techniques. This research presents KDD models (Knowledge Discovery in Databases) that are useful for an entity management.

From the theoretical point of view, the KDD process will be considered to create a predictive model to get a high level of effectiveness and the different class balancing techniques will be studied to improve data quality.

In this way, the organization expects to plan the new definitions on the active beneficiaries' treatment from a new focus, using new retention techniques which enable the senior managers to count with true and accurate information as a support to make the most suitable decisions. In conclusion, these models directly influence competitiveness and the planning process.

## Future Research Lines

There are some algorithms that were left aside of this research but which can be relevant for imbalance of classes, i.e., the logistic model, deep learning of hybrid resampling or the model confidence interval. Some of them had a more complex methodological development and a superior processing time than the ones studied; this is the reason why they were not included in this final work.

On the other hand, there is no evidence on this field in another section of the entity, so it will be relevant for the senior managers to carry out some research to clarify the specific role data mining techniques have within the companies. There exists a substantial motivation of future research within this field to get better results.

## ESPECIALIZACIÓN INTELIGENCIA DE DATOS ORIENTADA A BIG DATA



**Esp. Patricia Elizabeth Prado**

e-mail  
pprado@live.com.ar

español

**Director**  
Dr. Waldo Hasperué  
**Fecha de aprobación**  
11 de octubre de 2022  
**SEDICI**  
[Tesis Completa](#)

## Estudio de Técnicas de Agrupamiento en Procesos de Datos a Gran Escala

**Palabras clave:** Clustering; Spark; K-Means; Bisecting K-Means; Gaussian Mixture Model

### Motivación

En dominios donde el volumen de los datos crece exponencialmente y la enorme abundancia de estos desborda la capacidad humana para comprenderlos, existe una necesidad apremiante de desarrollar soluciones para aprovechar esta riqueza de datos utilizando métodos estadísticos.

En muchas situaciones, el método tradicional de convertir los datos en conocimiento consiste en un análisis e interpretación realizado en forma manual, mediante consultas efectuadas con lenguajes generalistas de consulta como SQL sobre la base de datos operacional o ejecutando operaciones de procesamiento analítico en línea sobre almacenes de datos. Estas soluciones admiten técnicas de análisis como pueden ser el resumen, la consolidación, la agregación y la posibilidad de ver la información desde distintas perspectivas. Sin embargo, no generan reglas, patrones, pautas, es decir conocimiento que pueda ser aplicado a otros datos.

El agrupamiento es la tarea descriptiva por excelencia, consiste en obtener grupos naturales a partir de los datos para luego poder describirlos de manera concisa. Ya sea para la comprensión o el resumen, el análisis de agrupamiento ha desempeñado durante mucho tiempo un papel importante en una amplia variedad de campos como psicología, ciencias sociales, biología, estadísticas, reconocimiento de patrones y recuperación de información.

En el marco de la pandemia del COVID-19, se dispone de información sobre millones de casos de contagios registrados en Argentina con sus características personales, geográficas y temporales. Resulta de interés segmentar los casos en grupos tales que las muestras de un grupo permitan extraer descripciones que resulten de utilidad para el sistema de salud, para este fin las técnicas de agrupamiento son una valiosa herramienta.

El presente estudio pretende analizar diferentes técnicas de agrupamiento en el entorno distribuido Spark, que permitan

describir de forma novedosa el seguimiento de casos de COVID-19 en Argentina a partir de la aplicación de modelos de agrupamiento adecuados para el desarrollo de aplicaciones de aprendizaje automático a gran escala.

### Aportes del Trabajo Final Integrador

Se describen los resultados de los ensayos realizados que consistieron en la ejecución de las técnicas de agrupamiento con diferentes configuraciones de sus hiperparámetros, utilizando diferentes tamaños del conjunto de datos para evaluar escalabilidad y distintos grupos de características para determinar cómo influye la dimensionalidad. Se realizaron experimentos donde se ejecutaron los algoritmos en una única computadora, como así también en un clúster de tres nodos. En estos ensayos se midieron y compararon tiempos de ejecución.

Por otro lado, los grupos encontrados fueron evaluados con métodos de validación interna para determinar la pureza de estos, al mismo tiempo que se utilizaron técnicas de visualización y acumulación de evidencia para la comparación de los diferentes grupos de datos obtenidos por los diferentes modelos.

### Líneas de I/D futuras

Como trabajo futuro se sugiere explorar diferentes medidas de distancia y otras técnicas de clustering como algoritmos basados en densidad además de reiterar los ensayos realizados en el conjunto de datos actual



english

**Advisor**  
Dr. Waldo Hasperué  
**Thesis approval date**  
October 11, 2022  
**SEDICI**  
[Full Text](#)

## **Study of Clustering Techniques in Large Scale Data Processing**

**Keywords:** Clustering; Spark; K-Means; Bisecting K-Means; Gaussian Mixture Model

### **Motivation**

In domains where the volume of data grows exponentially and the enormous abundance of data exceeds the human capacity to understand it, there is a pressing need to develop solutions to take advantage of this wealth of data using statistical methods.

In many situations, the traditional method of converting data into knowledge consists of manual analysis and interpretation, using queries performed with generalist query languages such as SQL on the operational database or by executing online analytical processing operations on data warehouses. These solutions support analysis techniques such as summary, consolidation, aggregation and the possibility of seeing information from different perspectives. However, they do not generate rules, patterns, guidelines, that is, knowledge that can be applied to other data.

Clustering is the descriptive task par excellence, it consists of obtaining natural groups from the data and then being able to describe them concisely. Whether for understanding or summary, clustering analysis has long played an important role in a wide variety of fields such as psychology, social sciences, biology, statistics, pattern recognition, and information retrieval.

In the context of the COVID-19 pandemic, information is available on millions of cases of infections registered in Argentina with their personal, geographical and temporal characteristics. It is of interest to segment the cases into groups such that the samples of a group allow to extract descriptions that are useful for the health system, for this purpose the grouping techniques are a valuable tool.

This study aims to analyze different clustering techniques in the Spark distributed environment, which allow to describe in a novel way the monitoring of COVID-19 cases in Argentina from the application of clustering models suitable for the development of large-scale machine learning applications.

### **Thesis Final Work contributions**

The results of the tests performed that consisted of the execution of clustering techniques with different configurations of their hyperparameters, using different sizes of the dataset to evaluate scalability and different groups of characteristics to determine how dimensionality influences are described. Experiments were conducted where the algorithms were run on a single computer, as well as on a three-node cluster. In these trials, execution times were measured and compared.

On the other hand, the groups found were evaluated with internal validation methods to determine the purity of these, at the same time that visualization and evidence accumulation techniques were used for the comparison of the different groups of data obtained by the different models.

### **Future Research Lines**

As future work, it is suggested to explore different distance measurements and other clustering techniques such as density-based algorithms in addition to reiterating the tests carried out on the current dataset.

ESPECIALIZACIÓN  
COMPUTACIÓN GRÁFICA,  
IMÁGENES Y VISIÓN  
POR COMPUTADOR



**Esp. Cicerchia Lucas Benjamin**

e-mail

lucas.cicerchia@itt.unnoba.edu.ar

**Director**  
Dra. María José Abásolo

**Codirector**  
Dra. Claudia Cecilia Russo

**Fecha de aprobación**  
1 de diciembre de 2022

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

## **Clasificación de cultivos en imágenes Landsat utilizando algoritmos de Active Learning**

**Palabras clave:** Machine Learning; Active Learning; Procesamiento de Imágenes; Sensado Remoto; Clasificación de Cultivos

### **Motivación**

El presente trabajo surge como una búsqueda de desarrollo de una herramienta que permita a instituciones o entidades ligadas a la actividad de la Agricultura de Precisión (AP) a obtener una clasificación de cultivos dentro de la región del noroeste de la provincia de Buenos Aires de manera simple y rápida. En base al estudio de cómo trabajaban entidades como el Instituto Nacional de Tecnología Agropecuaria Estación Experimental Agropecuaria Pergamino (INTA EEA Pergamino) en relación al relevamiento de cultivos que se sembraban en la región es que surge la idea de abordar esta temática, buscando por un lado objetivar el trabajo realizado, mejorar los resultados que estos obtienen, simplificar la tarea lo que permite un ahorro de tiempo y por último comenzar a incorporar otros cultivos que en el relevamiento que estos realizan no se identifican. Sumado a ello, al no contar con una descripción de los lotes de la región hace que la separación entre cultivos sea dificultosa para las técnicas de clasificación que utilizan, lo que lleva en muchas ocasiones a confusiones por parte del algoritmo.

Entre las tecnologías que utilizan se encuentra el Sensado Remoto (RS), que incluye imágenes de satélite, el cual genera una gran cantidad de datos para la AP los cuales requieren del uso de técnicas de Machine Learning (ML) para obtener información a partir de estos. Entre las diferentes técnicas de ML más recientes, se destacan algunas de las más utilizadas Semi-Supervised Learning (SSL), Support Vector Machine (SVM), Deep Learning (DL) y Active Learning (AL). En las técnicas como SVM o DL, todas las muestras o datos tienen que estar etiquetadas con la dificultad de que la gran cantidad de datos generados por los sensores remotos hoy en día, hace que el etiquetado de datos para los algoritmos sea una tarea difícil y que consume muchas horas de trabajo. Este problema hace que a menudo estos algoritmos pueden aplicarse a un conjunto de datos pequeño.

En contraparte a estas técnicas existen alternativas como el Active Learning (AL) que pretende obtener un rendimiento de clasificación satisfactorio con un menor número de muestras etiquetadas y que además explota la interacción continua con el clasificador con un menor costo de etiquetado. Esto permite un reentrenamiento constante del algoritmo, la explotación de los datos y reduce el coste del etiquetado. Las muestras a etiquetar se seleccionan mediante una heurística para que sean lo suficientemente representativas como para mejorar el modelo y luego serán etiquetadas por un experto. El presente trabajo tiene como objetivo explorar diferentes técnicas de AL aplicadas a RS en comparación con SVM para lograr un clasificador de diferentes cultivos de verano en el Noroeste de la provincia de Buenos Aires, Argentina. Más específicamente, la región de estudio que se ha utilizado comprende campos de los partidos de Junín, Rojas y Chacabuco, que forman parte de la llamada “Pampa Húmeda”, una de las regiones más relevantes en la producción agrícola y menos estudiada con este tipo de técnicas.

Lo que se considera relevante del objetivo, es por una parte la aplicación de este tipo de técnicas sobre una única imagen satelital del sensor Landsat. Por otra parte los tipos de cultivos que se estudian (Maíz, Soja y Sorgo) incluyendo además variaciones de un mismo cultivo.

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

Es decir que esta investigación surge a partir de la dificultad que estos tienen para poder realizar un relevamiento de los cultivos, y el área cultivada dentro de la región, donde las herramientas y técnicas utilizadas insumen mucho tiempo, en muchos casos tienen dificultades, y además emplean muchas imágenes a lo largo de la evolución del cultivo requiriendo de la utilización de imágenes con índices de vegetación. Donde para el abordaje del trabajo se utilizaron algunas de las tecno-



logías y técnicas mencionadas en el apartado anterior.

La particularidad de este desarrollo radica en cuatro aspectos fundamentales; en primer lugar, se trabaja con imágenes Landsat cuando muchos trabajos se realizan con imágenes hiperespectrales, multitemporales o multifuentes. Al trabajar con imágenes de este satélite, se dispone de una menor cantidad de datos que en otros casos que se utiliza imágenes hiperespectrales donde dispone de una resolución de 200 bandas espectrales que aportan información. En segundo lugar, para las pruebas de clasificación se utiliza una imagen de una sola fecha pero dentro de un rango de fechas determinado, que son las que los especialistas consideran que allí se pueden distinguir los cultivos a evaluar. La decisión de no utilizar imágenes multitemporales radica en que se busca verificar que se puede obtener un buen clasificador para estos cultivos con una sola imagen con el objetivo final de intentar simplificar la posterior utilización del clasificador por parte del usuario final. El tercer aspecto a destacar es que se aplica a una región donde este tipo de algoritmos no ha sido muy estudiado hasta la actualidad. Existen algunos trabajos sobre la cobertura en nuestro país, pero se aplican en otras regiones, a diferentes coberturas, utilizando imágenes multitemporales y con diferentes índices de vegetación. Y el último aspecto a tener en cuenta es que entre los cultivos a clasificar hay variedades de dos de los cultivos a clasificar. Es decir, se pretende discriminar el maíz en tres clases: Maíz de Primera, Maíz Tardío y Maíz de Segunda (de Doble Cultivo), y la soja en dos clases: Soja de Primera y Soja de Doble Cultivo (de Segunda), además de Sorgo y Agua. El principal reto es que son variantes de los mismos cultivos y tienen la misma firma espectral.

### Líneas de I/D futuras

El trabajo realizado hasta el momento es parte del trabajo realizado para una tesis doctoral para un Doctorado en Ciencias Informáticas de la Universidad Nacional de La Plata, en el cual se busca obtener una herramienta que permita obtener un clasificador de cultivos que sea rápido y simple de utilizar, a partir de este tipo de técnicas de AL aplicado a datos propios de coberturas e imágenes de RS de la región

El objetivo de las pruebas futuras por un lado, se espera realizar pruebas con más conjuntos de datos, incorporando imágenes de otros años con el fin de estudiar cómo las variaciones climáticas y los cambios en las fechas de siembra de los cultivos pueden afectar a la hora de obtener un buen clasificador. Además estudiar la posibilidad de incorporar más clases al conjunto de datos, añadiendo coberturas como girasol, suelo desnudo, bosques, construcciones o rutas.

En lo que a las técnicas de ML respecta se buscará continuar con el estudio de técnicas y heurísticas para lograr modificaciones que mejoren los resultados obtenidos hasta el momento. Se considera que la incorporación de técnicas de procesamiento digital de imágenes puede aportar a la mejora de estas técnicas.





**Advisor**  
Dra. María José Abásolo

**Codirector**  
Dra. Claudia Cecilia Russo

**Thesis approval date**  
December 1st, 2022

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

## **Crop classification on Landsat imagery using Active Learning algorithms**

**Keywords:** Machine Learning; Active Learning; Image Processing; Remote Sensing; Landcover Classification

### **Motivation**

The present work arises as a search for the development of a tool that allows institutions linked to the activity of Precision Agriculture (PA) to obtain a crop classification in the northwestern region of the province of Buenos Aires in a simple and fast way. Based on the study of how entities such as the National Institute of Agricultural Technology Pergamino Agricultural Experimental Station (INTA EEA Pergamino) work in relation to the survey of crops planted in the region, the idea of tackling this issue was started, seeking on the one hand to objectify the work, improve the results obtained, simplify the task which saves time and finally begin to incorporate other crops that are not identified in the survey carried out by them. In addition, not having a description of the plots in the region makes the separation between crops difficult for the classification techniques used, which often causes confusion in the algorithm.

One of the technologies used is Remote Sensing (RS), including satellite images, which generates a large amount of data for PA and requires the use of Machine Learning (ML) techniques to obtain information from them. Some of the most recent ML techniques are Semi-Supervised Learning (SSL), Support Vector Machine (SVM), Deep Learning (DL) and Active Learning (AL). In techniques such as SVM or DL, all samples or data have to be labeled with the difficulty that the large amount of data generated by remote sensors, makes labeling data for the algorithms a difficult task that consumes many hours of work. This problem often causes that these algorithms can be applied to a small dataset.

In contrast to these techniques there are alternatives such as Active Learning (AL) which aims to obtain successful

classification performance with a smaller number of labeled samples and also exploits continuous interaction with the classifier with a lower labeling cost. This allows constant training of the algorithm, data mining and reduces the cost of labeling. The samples to be labeled are selected by a heuristic to make them representative enough to improve the model and will then be labeled by an expert.

The present work aims to explore different AL techniques applied to RS in comparison with SVM to achieve a classifier of different summer crops in the Northwest of Buenos Aires province, Argentina. More specifically, the study region used comprises fields in the districts of Junín, Rojas and Chacabuco, which are part of the so-called “Pampa Húmeda”, one of the most relevant regions in agricultural production and less studied with this type of techniques.

What is considered relevant to the objective is, on the one hand, the application of this type of techniques on a single satellite image of the Landsat sensor. On the other hand, the types of crops studied (Corn, Soybean and Sorghum) including variations of the same crop.

### **Thesis Final Work contributions**

This research starts from the difficulty to survey the crops and the cultivated area within the region, where the tools and techniques used are extremely time-consuming. In many cases it is difficult, and in addition it uses many images throughout the evolution of the crop, requiring the use of images with vegetation indexes. Some of the technologies and techniques mentioned in the previous section were used

to approach the work. The particularity of this development is based on four fundamental aspects; in the first place, the work is carried out with Landsat images when many papers are carried out with hyperspectral, multitemporal or multisource images. When working with images from this satellite, a smaller amount of data is available than in other cases using hyperspectral images with 200 spectral bands that provide information. Secondly, an image of a single date is used but within a certain date range for the classification tests, which are the dates where the specialists consider that the crops to be evaluated can be distinguished. The decision not to use multitemporal images was made to verify that a good classifier for these crops can be obtained with a single image, with the final objective of trying to simplify the subsequent use of the classifier by the end user.

The third aspect to highlight is that it is applied to a region where this type of algorithms has not been widely studied until now. There are some works on coverage in our country, but they are applied in other regions, to different coverages, using multitemporal images and with different vegetation indices. And the last aspect to take into account is that among the crops to be classified there are varieties of two of the crops to be classified. That is to say, we intend to discriminate corn in three classes: Corn, Late Corn and Double Crop Corn, and soybean in two classes: Soybean and Double Crop Soybean, as well as Sorghum and Water. The main challenge is that they are variants of the same crops and have the same spectral signature.

### Future Research Lines

This research is part of the work carried out for a doctoral thesis for a PhD in Computer Science at the National University of La Plata, in which the aim is to obtain a tool that allows to obtain a quick and simple to use crop classifier, based on this type of AL techniques applied using own data from RS coverages and images of the region.

The objective of future tests on the one hand, is to test with more data sets, incorporating images from other years in order to study how climatic variations and changes in crop planting dates can affect to obtain a good classifier. Also study the possibility to add more classes to the dataset, by incorporating coverages such as sunflower, bare soil, forests, constructions or roads.

Regarding ML techniques, we will continue with the study of techniques and heuristics to modify them in order to improve the results obtained so far. It is considered that the incorporation of digital image processing techniques can contribute to the improvement of these techniques.

## ESPECIALIZACIÓN TECNOLOGÍA INFORMÁTICA APLICADA EN EDUCACIÓN



**Esp. María Inés Kessler**

e-mail

ikessler@fahce.unlp.edu.ar

**Directora**  
Dra. Cecilia Verónica Sanz

**Fecha de aprobación**  
30 de marzo de 2022

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

## Las estrategias de formación de usuarios mediadas por tecnología informática en bibliotecas de instituciones de educación superior

**Palabras clave:** Formación de usuarios; Educación superior; Educación a distancia; Blended Learning; Bibliotecas universitarias; Entornos virtuales de Enseñanza Aprendizaje; Universidades; Argentina

### Motivación

Las bibliotecas universitarias han desarrollado, con el paso del tiempo, diversas estrategias para facilitar el acceso a los recursos de información, tanto de su propia colección como remotos. Entre las estrategias planteadas se encuentran los programas de formación de usuarios, a través de los cuales se busca formar usuarios autónomos en el uso de los servicios y recursos de información.

Los continuos avances de las tecnologías de información y comunicación, además de la incorporación de nuevos servicios y recursos de información, hacen necesaria la adaptación de los programas de formación de usuarios al nuevo contexto en el que las bibliotecas universitarias están insertas.

Se deben facilitar al estudiante las herramientas necesarias, tanto para acceder a la información durante el desarrollo de su carrera, como para convertirlo en un profesional autosuficiente en la búsqueda y recuperación de ésta, enfrentando los permanentes desafíos de los avances tecnológicos.

Se deben considerar, además de los factores antes mencionados, las transformaciones producidas en las modalidades educativas que actualmente se ofrecen, para poder evaluar y definir una estrategia adecuada a las necesidades de formación que demandan los usuarios en la actualidad.

Esta situación, con los cambios descriptos en el entorno, pero además hacia el interior de las instituciones, plantea el interrogante de saber en qué situación se encuentran los programas de formación de usuarios en las bibliotecas universitarias, y motiva la necesidad de indagar sobre las estrategias de formación que se desarrollan desde estas instituciones, y especialmente aquellas estrategias mediadas por tecnología informática.

La investigación que se propone, a través del relevamiento de experiencias, y del análisis de las estrategias y modalidades empleadas en la formación de usuarios, proporciona un panorama general de la situación actual en cuanto a las ofertas de formación mediadas por tecnología informática de las bibliotecas universitarias de universidades públicas argentinas.

Se analiza lo que proyectan para sus usuarios, permitiendo detectar, además, aquellas que no ofrecen formación, obteniendo un porcentaje de los casos en ambas situaciones, que permite realizar una valoración de la situación actual que, además, se encuentra atravesada por lo ocurrido con la pandemia. Se espera que el trabajo sirva como motivación a aquellas instituciones que no poseen programas de formación de usuarios, y también, por qué no, a aquellas que sí cuentan con estos programas, pero deseen proponer algún tipo de innovación a futuro.

#### Como objetivo general de este trabajo se plantea

-Realizar un análisis teórico exploratorio sobre estrategias y modalidades de formación de usuarios de bibliotecas universitarias, de Argentina, que incorporen tecnologías de información y comunicación para la mediación.

#### Los objetivos específicos se orientan a:

- Investigar sobre modalidades educativas: a) según la tecnología aplicada (m-learning, t-learning, e-learning) y b) de acuerdo a las estrategias didácticas y de mediación de contenidos (blended learning, o extended learning);

- Realizar el relevamiento de experiencias y proyectos de formación de usuarios en bibliotecas universitarias (de universidades públicas), que apliquen tecnología informática en sus propuestas;

- Analizar y comparar los casos, resultado del relevamiento, a partir de criterios propios que permitan revisar aspectos educativos y tecnológicos de formación de las propuestas estudiadas.

### Aportes del Trabajo Final Integrador

Este trabajo, de tipo exploratorio, presenta, mediante la indagación teórica, la exploración y análisis de sitios web y la realización de encuestas, el estado actual de los programas de formación de usuarios, con foco en propuestas con mediación tecnológica, en las bibliotecas universitarias, de gestión pública, de Argentina.

El trabajo describe los aspectos generales de la formación de usuarios, su evolución y su implementación en bibliotecas universitarias. Aborda la temática de educación a distancia, planteando aspectos generales, su contextualización y modalidades. Se fundamenta la realización de este tipo de modalidades teniendo en cuenta su importancia para las instituciones en cuestión. Se plantea y desarrolla una propuesta de investigación a partir del relevamiento de información pertinente, sobre las prácticas de formación de usuarios, a través de los sitios web de bibliotecas universitarias y de encuestas a informantes clave, para determinar el estado actual del tema en el país.

Los resultados alcanzados en el trabajo permiten ver que existe un escaso número de bibliotecas que ofrecen propuestas de formación de usuarios mediadas por tecnología informática. La modalidad educativa que prevalece, según la frecuencia del encuentro, es la presencial, en menor medida se ofrecen programas semipresenciales y a distancia, destinados a usuarios alumnos en todos los casos, y a otros actores institucionales en menor medida. Si bien se valoran como positivas las experiencias existentes, es necesario considerar que las bibliotecas deben asumir un rol activo, con mayor fuerza, en el compromiso de acompañar el proceso de adquisición de habilidades y competencias para el aprovechamiento de los recursos, y para lograr usuarios críticos en el uso de la información. Una de las estrategias es, justamente, a partir de la oferta de formación, aprovechando las ventajas de los recursos tecnológicos disponibles a la hora de pensar estrategias para la implementación.

### Líneas de I/D futuras

Como profesionales de la información es necesario reconocer la responsabilidad, al mismo tiempo que la posibilidad, de fomentar el buen uso de recursos y servicios, de formar usuarios responsables, allí donde el conocimiento nace y se desarrolla, en la biblioteca universitaria, en la universidad, en el inicio de las búsquedas bibliográficas y en el uso ético de la información. Las experiencias de otras instituciones son muy valiosas, observar el trabajo de colegas puede agregar valor a nuestra propia experiencia, servir de guía, provocar nuevas inquietudes.

Se considera que es preciso avanzar en la investigación. Como desafío futuro, se plantea la necesidad de investigar y proponer la implementación de programas de formación de usuarios, mediados por tecnología informática, en las bibliotecas que no cuenten con éstos. Para lograr este desafío, es posible plantear el desarrollo de un proyecto, adaptable a cualquier biblioteca universitaria, en el que se cubran todas las necesidades de formación básicas.



**Advisor**  
Dra. Cecilia Verónica Sanz

**Thesis approval date**  
March 30, 2022

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

## **User training strategies delivered by means of computer technology in libraries embedded in higher education institutions**

**Keywords:** User training; Higher education; Remote learning; Blended Learning; University libraries; Virtual environments of learning/teaching; Universities; Argentina

### **Motivation**

In order to simplify the access to information resources, such as universities' own collection or remote resources, university libraries have developed several strategies over time. One of these planned strategies is the development of user training programs, whose purpose is to train independent users for the handling of information services and resources, among others. The constant strides of information and communications technology, together with the development of new information services and resources, pushes the adaptation of user training programs in order to fit in with the new context of university libraries.

Students have to receive the necessary tools to access information. This includes tools to access the information needed during their course of studies, as well as the tools that improve their ability to self-search and successfully retrieve information at their professional level, all in the face of the repeated challenges of technological breakthroughs. Another aspect to consider is the transformation of the educational modalities, which affects the evaluation and establishment of an appropriate strategy to satisfy the training needs of users nowadays.

Taking into consideration all of the above mentioned changes in the environment, and in the internal context of the institutions, a question arises regarding the current state of user training programs in university libraries and pushes the need of investigating the training strategies developed in these institutions, especially those related to computer technology. Through a survey of experiences and an analysis of the strategies and modalities implemented in user training, the proposed research provides a broad overview of the current state of training options delivered by means of computer technology in public university libraries in Argentina. The analysis screens their projects available for users; this also helps identify which university libraries do not offer training options. This analysis yields a percentage of both

scenarios and provides the basis for an assessment of the current situation in a post pandemic world.

This work is intended to serve as a motivation for the institutions that do not own user training programs and for those that might want to improve their current programs in the future.

#### **The general purpose of this work is:**

- To perform an exploratory theoretical analysis of user training programs in university libraries in Argentina delivered by means of computer technology.

#### **The specific purposes of this work are:**

- To perform an investigation on the learning modalities according to a) applied technology (m-learning, t-learning, e-learning) and b) the teaching strategies and strategies to deliver content (blended learning or extended learning).
- To perform a survey of user experiences and user training projects of university libraries (public universities) delivered by means of computer technology
- To analyze and compare said cases and the results yield by the survey, following individual criteria that allows the assessment of educational and technological aspects of the training proposals present in this study.

### **Thesis Final Work contributions**

This exploratory study portrays, through theoretical research, investigation and analysis of websites, and surveys, the current state of user training programs in public university libraries in Argentina, with an emphasis on the technological means of such programs.

This work describes general aspects of user training, its development and implementation in university libraries. It addresses remote learning, including a brief description of this process, its context and modalities. We also analyze the importance of said modalities for the institutions studied in this paper.

We state and develop a line of investigation by reviewing relevant information related to user training practices in order to determine the current state of the subject in the country. The revision is done on the websites of the university libraries and through surveys to key informants.

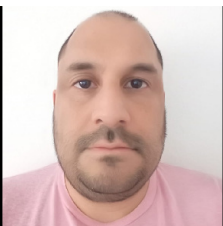
The results obtained by this work show us that there is a very low number of libraries that offer user training proposals delivered by means of computer technology. The main learning modality, according to the frequency of meetings, is on-site learning. Then, to a smaller extent, there are blended learning and remote learning programs, meant for students in most cases and for other participants of the site on a lower scale. While existing experiences are considered positive, it is important for libraries to take a stronger active role and commitment in supporting the process of acquiring skills and abilities to leverage resources; and to train users for the conscious use of information. Precisely, a way to do this is to provide training options, leveraging the technological resources available when planning strategies to implement.

### Future Research Lines

As information professionals, it is necessary to acknowledge our responsibility, and possibility, to encourage the correct use of resources and services, and to train responsible users in their literature searches and ethical use of information, specially in the context of a university library, where knowledge rises and develops. The experiences of other institutions are very meaningful, as well as observing other colleagues' work, to add value to our own experience, guide our efforts and pose new questions.

It is of paramount importance to continue this line of investigation. As a future challenge, there is a need to investigate and project the implementation of user training programs delivered by means of computer technology in the libraries where this service is not available. To achieve this goal, it is possible to plan and develop a project that covers all those training requirements, and a project that can be adapted to the university library's needs.

## ESPECIALIZACIÓN TECNOLOGÍA INFORMÁTICA APLICADA EN EDUCACIÓN



**Esp. Daniel Antonio  
Arellano Sarmiento**

e-mail  
aredan25@gmail.com

español

**Director**  
Mg. Gustavo Astudillo

**Codirectora**  
Mg. Fernanda Barranquero

**Fecha de aprobación**  
27 de mayo de 2022

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

## **Un estado del arte y análisis de experiencias de uso de la robótica educativa en la escuela primaria pública de Chile, Uruguay y Argentina**

**Palabras clave:** Robótica; Robótica educativa; Pensamiento Computacional

### **Motivación**

Uno de los grandes retos para la educación del siglo XXI lo constituyen las tecnologías digitales; en especial se ha hecho énfasis en el desarrollo del Pensamiento Computacional (PC), que, si bien tiene sus raíces en las ideas de Seymour Papert en los '70, es retomado y desarrollado por Jeannette Wing a inicios del 2006.

Se puede mencionar que, una de las estrategias utilizadas para desarrollar el PC es la robótica educativa (RE) o también llamada robótica pedagógica que Ruiz-Velasco (2017) define como una disciplina que permite concebir, diseñar y desarrollar robots educativos para que los alumnos se inicien desde muy jóvenes en el estudio de las ciencias y la tecnología.

Por tal motivo, es que se propone la realización de este trabajo que tiene como objetivo general el Analizar experiencias de uso de la robótica educativa y las bases teóricas que las sustentan, en la escuela primaria pública de Chile, Uruguay y Argentina.

#### **Como objetivos específicos se plantean:**

- Identificar las bases teóricas y/o enfoques que sustentan la robótica educativa. Identificar las distintas experiencias de implementación realizadas, en Chile, Uruguay y Argentina, que estén orientados al uso de la Robótica Educativa en el nivel primario de educación.
- Definir un conjunto de criterios de análisis de las propuestas didácticas que subyacen en las experiencias identificadas
- Analizar las experiencias con base en los criterios propuestos en función de identificar fortalezas y debilidades.
- Proponer el desarrollo de futuras líneas de investigación relacionadas con el uso de la robótica educativa en la escuela primaria.

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

Este trabajo se ha realizado una revisión bibliográfica sobre las bases teóricas relacionadas con el pensamiento computacional, la robótica y la robótica educativa. Además, se identificaron un conjunto de experiencias educativas a las que se les aplicó una serie de criterios para su inclusión o exclusión. Posteriormente, las experiencias seleccionadas, fueron sometidas a un proceso de análisis donde se pudieron identificar sus fortalezas y sus debilidades. Finalmente, se propusieron futuras líneas de investigación relacionadas con el uso de la robótica en la escuela primaria.

### **Líneas de I/D futuras**

En cuanto a las futuras líneas de investigación, se pretende profundizar sobre esta temática, a través de la definición de un conjunto de criterios que permitan clasificar y analizar las distintas propuestas didácticas que incluyen RE dentro de su estrategia de enseñanza, determinar metodologías de enseñanza que incluyan RE y posibles impactos esperados y no esperados en el proceso de aprendizaje. De esta forma poder medir la influencia que la RE puede tener en los alumnos y alumnas en su aprendizaje.





english

**Advisor**  
Mg. Gustavo Astudillo

**Codirector**  
Mg. Fernanda Barranquero

**Thesis approval date**  
May 27, 2022

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

## **Un estado del arte y análisis de experiencias de uso de la robótica educativa en la escuela primaria pública de Chile, Uruguay y Argentina**

**Keywords:** Robotics; Educational Robotics; Computational Thinking

### **Motivation**

One of the great challenges for education in the 21st century is digital technologies; In particular, emphasis has been placed on the development of Computational Thinking (CT), which, although it has its roots in the ideas of Seymour Papert in the '70s, is revisited and expanded on by Jeannette Wing in early 2006.

It can be mentioned that one of the strategies used to develop CT is educational robotics (ER). Also known as pedagogical robotics, which is defined by Ruiz-Velasco (2017) as a discipline that allows conceiving, designing and developing educational robots so that students embark upon at an early age the study of science and technology.

Hence the proposal of this study, which has as its main objective to, analyze experiences of use of educational robotics and the theoretical bases that support them, in the public primary school of Chile, Uruguay and Argentina.

As specific objectives: Identify the theoretical bases and / or approaches that support educational robotics. Identify the different implementation experiences carried out, in Chile, Uruguay and Argentina, that are oriented to the use of Educational Robotics at the primary level of education. Define a set of criteria for the analysis of the didactic proposals that underlie the identified experiences. Analyze the experiences based on the proposed criteria in order to identify strengths and weaknesses. Propose the development of future lines of research related to the use of educational robotics in primary school.

### **Thesis Final Work contributions**

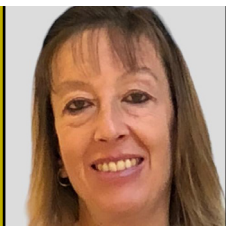
This paper is on a bibliographic review on the theoretical bases related to computational thinking, robotics and educational robotics. In addition, a set of educational experiences were identified to which a series of selection criteria were applied to define their inclusion or exclusion. Subsequently, the selected experiences were subjected to an analysis process where their strengths and weaknesses could be identified. Finally, future lines of research related to the use of robotics in primary school were proposed.

### **Future Research Lines**

Regarding future lines of research, it is intended to study in more detail this subject, through the definition of a set of selection criteria that allows classifying and analyzing the different didactic proposals that include RE within their teaching strategy, determining teaching methodologies that include RE and possible expected and unexpected impacts on the learning process. To this extent weigh RE influence on students learning process.



## ESPECIALIZACIÓN TECNOLOGÍA INFORMÁTICA APLICADA EN EDUCACIÓN



**Esp. María Cecilia Pezzini**

e-mail

c\_pezzini@hotmail.com

**Director**  
Mg.Pablo Thomas

**Fecha de aprobación**  
19 de octubre de 2022

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

## **Tutores inteligentes en la enseñanza: una revisión y análisis en la educación secundaria**

**Palabras clave:** Sistemas tutores inteligentes; Educación personalizada; Sistemas inteligentes

### **Motivación**

De manera periódica, Argentina participa en estudios y programas regionales e internacionales de evaluación de aprendizajes, como complemento de las evaluaciones que se desarrollan anualmente en el país, por parte de la Secretaría de Evaluación Educativa del Ministerio de Educación, Cultura, Ciencia y Tecnología de la Nación. Dentro de ellos se encuentra el Programa para la Evaluación Internacional de Estudiantes (Pruebas PISA por su sigla en inglés), implementado por la Organización para la Cooperación y el Desarrollo Económico (OCDE).

La Prueba PISA es un estudio internacional dirigido a estudiantes que tengan 15 años de edad y estén cursando 7° año o más. El operativo es trienal y permite evaluar los aprendizajes de los estudiantes en diferentes áreas del conocimiento y aporta, a su vez, referencias sobre los sistemas educativos de los países.

Los resultados de Argentina en las pruebas PISA 2018 son poco alentadores.

Argentina es el país de la región que más retrocedió con respecto a la última edición (2012): pasó de 388 puntos a 379. Se ubica por debajo de la media de América Latina. Los datos muestran que Argentina queda rezagada con respecto a otros países latinoamericanos, quedando en el puesto 8° de la región, debajo de Uruguay (418), Chile (417), México (409), Costa Rica (402), Perú (400), Colombia (391) y Brasil (384). Los estudiantes argentinos sólo superan a los de Panamá (353) y República Dominicana (325).

Argentina se ubica en el puesto 71 sobre 78 participantes.

Un modo de ayudar a fortalecer los procesos de aprendizaje, es mediante la utilización de sistemas tutores inteligentes. Se pondrá especial atención en aquellos Sistemas Tutores Inteligentes (en adelante STI), relacionados con la enseñanza de matemática. La preferencia se debe a que el pensamiento matemático es aquel que mayor dificultad presenta.

La mejora en el mismo, brinda herramientas para desarrollar el pensamiento analítico y el aprendizaje de otras disciplinas. Favoreciendo además, el desarrollo de capacidades de razonamiento y abstracción.

#### **Objetivo Generales**

Este trabajo tiene como objetivo estudiar y analizar el uso de STI, las metodologías que se aplican, y qué trabajos se realizan al respecto en la educación secundaria.

#### **Objetivos Específicos**

- Revisar el concepto de STI: componentes modulares (arquitectura de los STI), características.
- Estudiar qué tipos de trabajos de aplicación de STI se aplican en la educación secundaria.
- Identificar en las investigaciones revisadas:
  - La evolución y el estado de los desarrollos de STI, sistematizando el conocimiento del tema, en el contexto de la educación secundaria.
  - Las técnicas de personalización que permitan definir patrones de comportamiento de los estudiantes.
  - Las técnicas y métodos utilizados en la medición del efecto del uso de los STI.
  - Las debilidades encontradas en la implementación de los STI, en el marco de la educación secundaria.

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

Se definieron cinco aspectos a considerar al revisar los STI: aspectos generales, aspectos relacionados con el feedback del STI, aspectos metodológico-educativos, y de framework del STI. Del análisis de los papers, se desprenden las tendencias hacia las cuales avanzan los desarrollos de los STI. Estas tendencias involucran:

- La incorporación de agentes informáticos conversacionales, ayudando a la tutoría electrónica, a través del diálogo.
- El diagnóstico cognitivo, para determinar el nivel del estu-

dian­te en la ma­te­ria que se está tra­ba­jan­do, ge­ne­ran­do una ru­ta de apren­di­za­je per­so­na­li­za­da, de acuer­do a las ne­ce­si­da­des del alu­m­no.

- La in­te­gra­ción de STI con for­ta­le­zas com­ple­men­ta­rias para po­ten­ciar el apren­di­za­je y per­mitir la com­ple­men­ta­riedad de los re­cur­sos de apren­di­za­je. Es el ca­so de Shareable Knowledge Objects as Portable Intelligent Tutors (SPOKE IT, por sus siglas en inglés) , y MathSpring.

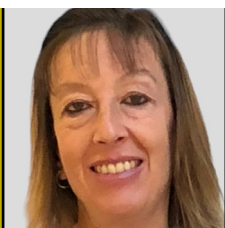
Los STI tie­nen al es­tu­dian­te como el cen­tro del pro­ce­so edu­ca­tivo, sien­do éste quien re­gula su apren­di­za­je. Los há­bi­tos de es­tu­dio au­to­re­gula­dos, se trans­for­man en­ton­ces, en un ele­men­to de­ter­mi­nan­te para el éxi­to del pro­ce­so edu­ca­tivo. El uso del STI, como he­rra­mien­ta com­ple­men­ta­ria, re­duce las di­fe­ren­cias en­tre alu­m­nos de un mis­mo curso de ma­te­má­ti­ca, al tie­mpo que da lu­ces a los docen­tes sobre el es­ta­do de apren­di­za­je de cada uno de sus alu­m­nos, per­mitién­do­les lle­var un con­trol de­ta­lla­do de sus di­fi­cul­ta­des y cono­ci­mien­tos al­can­za­dos.

Las e­va­lua­cio­nes de los STI, lle­va­das a ca­bo por los au­to­res de los tra­ba­jos mue­stran que los sis­te­mas de tu­to­ría, su­peran a los tu­to­res no ex­per­tos e in­clu­so po­drían igua­lar a los tu­to­res hu­ma­nos ex­per­tos, en al­gunos te­mas.

### **Líneas de I/D futuras**

Los re­sul­ta­dos de este tra­ba­jo son un pun­to de par­ti­da para con­ti­nuar con el de­sa­fío de me­jo­rar el apren­di­za­je de los es­tu­dian­tes en el área de ma­te­má­ti­ca, don­de se han agra­va­do aún más las de­fi­ci­en­cias en el apren­di­za­je, como con­se­cuen­cia de la pan­de­mia del COVID 19.

Como tra­ba­jo fu­tu­ro se plan­tea im­ple­men­tar el uso de al­guno de los STI ana­li­za­dos en alu­m­nos de 1º año de edu­ca­ción se­cun­da­ria y es­tu­diar su im­pac­to en el apren­di­za­je.



**Advisor**  
Mg.Pablo Thomas  
**Thesis approval date**  
October 19, 2022  
**SEDICI**  
[Full Text](#)

## **Intelligent Tutoring in Teaching: A Review and Analysis in Secondary Education**

**Keywords:** Intelligent tutoring systems; Customized education; Intelligent systems

### **Motivation**

Argentina regularly participates in regional and international learning assessment studies and programs, as a complement to the evaluations carried out every year in the country by the Secretary of Educational Evaluation of the National Ministry of Education, Culture, Science and Technology. Among these, the Program for International Student Assessment (known as PISA testing) is implemented by the Organization for Economic Co-operation and Development (OECD).

The PISA test is an international assessment intended for students aged 15 years who are enrolled in 7th grade or more. It is triennial and it allows for the evaluation of students' learning in various areas of knowledge, while providing references about the educational system of different countries.

The results of Argentina in PISA tests in 2018 were not very encouraging. Argentina is the country in the region which showed the biggest regress with respect to the previous run (2012): it fell from 388 points to 379. It is below average in Latin America. Data shows that Argentina lags behind other Latin American countries, and is ranked eighth in the region, after Uruguay (418), Chile (417), Mexico (409), Costa Rica (402), Peru (400), Colombia (391), and Brazil (384). Argentine students only outperform students from Panama (353) and Dominican Republic (325).

Argentina is ranked 71st over 78 participants.

The use of intelligent tutoring systems can serve as a way of helping to strengthen learning processes. Special emphasis shall be put on Intelligent Tutoring Systems (hereinafter, ITS) in relation to the teaching of Math. This preference is based on the fact that the mathematical thinking is that which presents the greatest difficulty. Improving it will provide certain tools to develop analytical thinking and learn content from other disciplines, while promoting an advancement in reasoning and abstraction skills.

### **General Objective**

This work aims at studying and analyzing the use of ITS, the methodologies applied and the related works carried out in secondary education.

### **Specific Objectives**

- Review the concept of ITS: modular components (the architecture of ITS), characteristics.
- Study what types of ITS-implementation works are applied in secondary education.
- In the investigations under scope, identify:
  - The evolution and state of ITS developments, systematizing the knowledge of the subject matter within the context of secondary education.
  - Customization techniques enabling the identification of students' behavior patterns.
  - The techniques and methods used to measure the effect of ITS use.
  - The weaknesses found in the implementation of ITS within the framework of secondary education.

### **Thesis Final Work contributions**

Five aspects were defined for consideration upon the review of ITS: general aspects, aspects related to ITS feedback, methodological-educational aspects, and ITS framework.

The tendencies towards which ITS developments are progressing became clear after the analysis of literature. These tendencies involve:

- Incorporation of conversational computer agents to help electronic tutoring through dialogue.
- Cognitive diagnostics to determine the student's level in the subject under study, generating a customized learning path according to the learner's needs.
- Integration of ITS with complementary strengths to enhance learning and enable the complementarity of learning resources. For instance, Shareable Knowledge Objects as

Portable Intelligent Tutors (SPOKE IT) and MathSpring.

In ITS the student is at the core of the educational process by regulating his/her own learning.

Self-regulated study habits are then transformed into a determining factor for a successful educational process.

The use of ITS as a complementary tool reduces the differences among students from the same Math course and enlightens teachers about the degree of learning of every student and allows them to keep track of the difficulties and knowledge attained by each student.

The ITS evaluations carried out by the work writers show that tutoring systems outperform non-expert tutors and may even match up to human expert tutors in some topics

### **Future Research Lines**

The results of this work serve as a starting point to continue with the challenge of improving students' learning in the area of Math, in which deficiencies have been further aggravated as a result of the COVID-19 pandemic.

As part of our future work, the aim is to implement the use of some of the ITS under study in students enrolled in 1st grade of secondary education and analyze the impact of these ITS on learning.

## ESPECIALIZACIÓN REDES Y SEGURIDAD



**Esp. Fernando Ariel Boettner**

e-mail  
fboettner@gmail.com

español

**Director**  
Dr. Fernando G. Tinetti

**Fecha de aprobación**  
5 de diciembre de 2022

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

## Mejoras en la seguridad web del usuario mediante el uso de un proxy local

**Palabras clave:** Cookies; Proxy; Sesión; Web; Seguridad; Identidad

### Motivación

Como es sabido, HTTP es un protocolo que no maneja estados. Esto significa que cada conexión que genera un usuario desde un navegador hacia un servidor web es independiente. La técnica utilizada para que los servidores web puedan tener trazabilidad de las acciones que realizan los usuarios es el uso de cookies. De hecho, gracias a esta técnica los servidores web logran mantener sesiones de usuario, a pesar de que cada petición proveniente del navegador web sea independiente de la anterior.

Si bien esta tecnología funciona correctamente, la privacidad de la navegación puede verse comprometida por el robo de las cookies. Existen distintos mecanismos que permiten a los atacantes obtener las cookies que se intercambian entre navegadores y servidores web. Por un lado, técnicas basadas en la obtención de cookies mediante escucha o sniffing de la red, y por el otro, mecanismos para obtener dichas cookies directamente desde los navegadores. En ambos casos, los atacantes pueden lograr la suplantación de identidad, robo de sesiones o incluso denegaciones de servicio.

El objetivo de este trabajo es analizar el uso de un proxy local como herramienta utilizada para mitigar algunos ataques a los que usuarios de aplicaciones web están expuestos cuando las mismas tienen vulnerabilidades que pueden permitir el robo y/o suplantación de identidad. Mediante el uso de un proxy local podría ser posible la manipulación de las cookies intercambiadas con el usuario de tal forma que los ataques a los que la aplicación podría estar expuesta no afecten la sesión del usuario. A los efectos de alcanzar esta meta se han definido los siguientes objetivos específicos:

1. Presentar los conceptos fundamentales del uso de cookies como mecanismo para mantener sesiones sobre el protocolo HTTP y HTTPS.
2. Exponer las amenazas más frecuentes a las que los usuarios están expuestos ante fallas en el manejo de sesiones de una aplicación web.
3. Analizar las técnicas existentes para mitigar dichas amenazas.
4. Proponer un método para evitar el robo o suplantación de sesiones, a partir de la utilización de un servicio de proxy local.

### Aportes del Trabajo Final Integrador

La utilización de cookies para el manejo de sesiones web se extendió muy rápidamente, en particular desde la llegada de la denominada Web 2.0. Con el paso de los años, se fueron actualizando y mejorando las versiones de los protocolos que transportan la información entre servidores y navegadores web, sin embargo, el uso de cookies sigue estando vigente.

El robo de cookies es una problemática compleja, ya que un atacante que sea capaz de obtener una cookie de sesión de un usuario, podrá utilizarla y entonces suplantar su identidad. La presente propuesta, no evita que una cookie sea robada, pero la manipulación de la misma “en tránsito”, mediante el proxy local, permite que esas cookies almacenadas en los navegadores web no posean los valores reales esperados por los servidores web, y de este modo, dejan de tener sentido la explotación de técnicas para obtener las cookies con ataques de tipo XSS.

### Líneas de I/D futuras

En un trabajo posterior, se pretende realizar el desarrollo de un software que, basado en el proxy local presentado aquí, permita realizar el intercambio de cookies, en forma automática y transparente para el usuario, almacenando internamente la relación “cookie real - cookie ficticia”. Adicionalmente, se podría avanzar en trabajos futuros sobre las siguientes temáticas relacionadas:

- Evaluación del mecanismo propuesto para el caso de cookies que no manejan sesiones, como cookies de personalización de contenidos y/o cookies de terceros.
- Manejo de sesiones utilizando Websockets.
- Análisis de seguridad de las sesiones de usuarios, cuando se utiliza JSON Web Tokens como técnica de autenticación.



**Advisor**  
Dr. Fernando G. Tinetti

**Thesis approval date**  
Decemeber 5, 2022

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

## Improvements in user web security through the use of a local proxy

**Keywords:** Cookies; Proxy; Session; Web; Security; Identity

### Motivation

It is well known that HTTP is a stateless protocol. This means that each connection generated by a user from a browser to a web server is independent. The use of cookies is the technique used for web servers to have traceability of the actions performed by users. In fact, thanks to this technique, web servers can maintain user sessions, even though each request from the web browser is independent of the previous one.

Although this technology works correctly, browsing privacy can be compromised by cookies theft. There are different mechanisms that allow attackers to acquire the cookies exchanged between browsers and web servers. On the one hand, techniques based on obtaining cookies by sniffing the network, and, on the other hand, mechanisms for obtaining those cookies directly from browsers. In both cases, attackers can achieve impersonation, session theft or even denial of service.

The objective of this work is to analyse the use of a local proxy as a tool to mitigate some attacks to which users of web applications are exposed when these applications have vulnerabilities that may lead to identity theft and/or impersonation. By using a local proxy, the manipulation of the cookies exchanged with the user could be possible in such a way that the attacks to which the application could be exposed do not affect the user's session. To achieve this goal, the following specific objectives have been defined:

1. To present the fundamental concepts of the use of cookies as a mechanism to maintain sessions over the HTTP and HTTPs protocols.
2. To expose the most frequent threats to which users are exposed, when facing faults in the management of web application sessions.
3. To analyse the existing techniques to mitigate such threats.
4. To propose a method to avoid session theft or impersonation, based on the use of a local proxy service.

### Thesis Final Work contributions

The use of cookies to manage web sessions spread very quickly, particularly since the arrival of the so-called Web 2.0. Over the years, the protocol versions that transport information between servers and web browsers have been updated and improved, but the use of cookies is still in force. Cookies theft is a complex problem, since an attacker capable of obtaining a user's session cookie can use it and, therefore, impersonate the user's identity.

The present proposal does not prevent a cookie from being stolen, but the manipulation of the cookie "in transit" through the local proxy allows that those cookies stored in web browsers do not have the real values expected by the web servers. Consequently, the exploitation of techniques to acquire cookies by means of XSS attacks does no longer make sense.

### Future Research Lines

It is considered for further work the development of a software that -based on the local proxy presented here- allows the automatic and transparent exchange of cookies for the user, internally storing the relationship "real cookie-fake cookie".

Additionally, future headway on the following related topics could be made:

- Evaluation of the proposed mechanism for cookies that do not handle sessions, such as content personalization cookies and/or third-party cookies.
  - Session handling using Websockets.
  - Security analysis of user sessions when JSON Web Tokens are used as authentication technique.
- on the learning process. To this extent weigh RE influence on students learning process.





# JURADOS DESIGNADOS 2022

Dra. Abasolo María José (UNLP - Argentina)  
Dr. Álvarez Hamelin José Ignacio (UBA - Argentina)  
Dr. Antonelli Leandro (UNLP - Argentina)  
Azrilevich Paola (MINCYT - Argentina)  
Mg. Banchoff Claudia (UNLP - Argentina)  
Dr. Balatti Pedro (UNLP - Argentina)  
Dra. Baldassarri Sandra (UNIZAR - España)  
Dr. Ballardini Javier (UNCOMA - Argentina)  
Dra. Bazán Patricia (UNLP - Argentina)  
Mg. Bertone Rodolfo (UNLP - Argentina)  
Mg. Bibbo Luis Mariano (UNLP - Argentina)  
Mg. Bria Oscar (UNLP - Argentina)  
Dr. Castro Antonio Lechtaler (UBA - Argentina)  
Dra. Cechich Alejandra (UNCOMA - Argentina)  
Esp. Delía Lisandro (UNLP - Argentina)  
Ing. De Giusti Armando Eduardo (UNLP - Argentina)  
Dra. De Giusti Laura (UNLP - Argentina)  
Dra. De Giusti Marisa (SEDICI - UNLP - Argentina)  
Dr. Errecalde Marcelo (UNSL - Argentina)  
Estrebou César (UNLP - Argentina)  
Dra. Estévez Elsa (UNLP - UNS - Argentina)  
Dr. Fernández Alejandro (UNLP - Argentina)  
Dr. Fillotrani Pablo (UNS - Argentina)  
Dr. Frati Emmanuel (UNDEC - Argentina)  
Dra. Gagliardi Edilma Olinda (UNSL - Argentina)  
Dr. García Garino Carlos (UNCuyo - Argentina)  
Dr. García Rubio Felix Oscar (UCLM - España)  
Dr. García Salvador (UGR - España)  
Dra. Genero Bocco Marcela (UCLM - España)  
Dra. Giandini Roxana (UNLP - Argentina)  
Dra. Godoy Daniela (UNCPBA - Argentina)  
Mg. González Alejandro (UNLP - Argentina)  
Dr. Guerrero Roberto (UNSL - Argentina)  
Dra. Guisen Andrea (UNR - Argentina)  
Mg. Harari Ivana (UNLP - Argentina)  
Dra. Herrera Susana (UNSE - Argentina)  
Dr. Jordan Ramiro (UNM - México)  
Dra. Lanzarini Laura (UNLP - Argentina)  
Dr. Leguizamón Guillermo (UNSL - Argentina)  
Mg. Llitteras Alejandra (UNLP - Argentina)  
Dra. Manresa Cristina (UIB - España)  
Ing. Marrone Luis (UNLP - Argentina)  
Dra. Mon Alicia (ITBA - Argentina)  
Dr. Navarro Martín Antonio (UCM - España)  
Dr. Olivas Varela José Ángel (UCLM - España)  
Dr. Olsina Luis (UNLPam - Argentina)  
Mg. Pérez Viviana (UBA - UNLP - Argentina)  
Dr. Piattini Mario (UCLM - España)  
Dra. Pons Claudia (UNLP - Argentina)  
Dr. Pousa Adrian (UNLP - Argentina)  
Dra. Printista Marcela (UNSL - Argentina)  
Lic. Queiruga Claudia (UNLP - Argentina)  
Dra. Reyes Nora (UNSL - Argentina)  
Dr. Riesco Daniel (UNSL - Argentina)  
Dr. Ronchetti Franco (UNLP - Argentina)  
Dr. Rosete Alejandro (CUJAE - Cuba)  
Dr. Rossi Gustavo (UNLP - Argentina)  
Dra. Russo Claudia (UNNOBA - Argentina)  
Dr. Salazar Chacón Gustavo David (ESPE - Chile)  
Dra. Sanz Cecilia Verónica (UNLP - Argentina)  
Dr. Suppi Boldrito Remo (UAB - España)  
Mg. Tugnarelli Mónica (UNER - Argentina)  
MSc. Velázquez Isaac Triguero (Un. Nottingham - UK)  
Mg. Venosa Paula (UNLP - Argentina)  
Dr. Villalba Luis Javier (UCM - España)  
Dra. Zangara María Alejandra (UNLP - Argentina)  
Dr. Zarza Gonzalo (Globant)  
Dr. Zunino Alejandro (UNCPBA - Argentina)

## INFORMES

<http://postgrado.info.unlp.edu.ar>

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

Tel/Fax: 54 221- 4273235

ISBN 978-950-34-2227-4



9 789503 422274

 postinfoUNLP

 postinfoUNLP

 postgradoinformaticaUNLP

 Postgrado Informática UNLP