

# Tesis Expuestas



Hemos seleccionados las últimas 10 Tesis de Doctorado y 5 de Maestría que se han defendido, de modo de tener una mirada rápida al conocimiento e innovación generados en las mismas. Es de hacer notar el crecimiento del número de Tesis de Doctorado defendidas, que es muy significativo. La riqueza temática de las Tesis de Doctorado queda reflejada en un rápido análisis del alcance de los 10 resúmenes que se presentan:

• **2 Tesis Doctorales relacionadas con Calidad, Datos Abiertos y Gobierno electrónico:**

- "Métricas de Calidad para validar conjuntos de datos abiertos públicos gubernamentales"
- "Modelos de Madurez para servicios de gobierno electrónico en el ámbito universitario"

• **4 Tesis Doctorales relacionadas con Ingeniería de Software:**

- "Modelado de Sistemas Colaborativos"
- "Formulación de procesos para una Ingeniería de Información Espacial"
- "DesProMod: Modelo de proceso de despliegue de sistemas de software"
- "Proceso de Requisitos Validado Empíricamente".

• **2 Tesis Doctorales relacionadas con Sistemas Inteligentes:**

- "Análisis y diseño de técnicas de preprocesamiento de instancias escalables para problemas no balanceados en Big Data. Aplicaciones en situaciones de emergencias humanitarias"
- "Generalización del modelado de cadencias de tecleo con contextos finitos para su utilización en ataques de presentación y canal lateral".

• **2 Tesis Doctorales relacionadas con Redes y Seguridad:**

- "Cálculo científico distribuido para clientes móviles indeterminados en redes MANETs.
- "Generalización del modelado de cadencias de tecleo con contextos finitos para su utilización en ataques de presentación y canal lateral"

Las Tesis de Maestría muestran diferentes perfiles que se alcanzan en el Postgrado. Un rápido análisis de los 5 resúmenes presentados se sintetiza a continuación.

• **3 Tesis de la Maestría en Ingeniería de Software:**

- "Arquitectura de Software en Wallet de código abierto para privacy coin en dispositivos móviles. El caso de estudio de Zcash".
- "Modelos de Predicción Avanzados para el cálculo de Reservas en la Industria Aseguradora"
- "Detección y Clasificación de Enfermedades en el Tomate Mediante Deep Learning y Computer Vision"

• **2 Tesis de la Maestría en Tecnología Informática aplicada en Educación:**

- "Estrategias de utilización de un simulador de entrevistas de trabajo para adultos hipoacúsicos con ayudas auditivas"
- "Realidad Aumentada en Contextos Educativos y su Relación con el Rendimiento Académico Universitario"

## Tesis en este número

### Doctorado en Ciencias Informáticas

#### 2022

María Roxana Martínez  
Ariel Pasini  
Jonathan Martin  
María José Basgall  
Luis Mariano Bibbo  
Giovanni Daián Rottoli  
Nahuel González  
Pablo José Iuliano  
Marisa Daniela Panizzi  
Gladys N. Kaplan

### Maestría

#### INGENIERÍA DE SOFTWARE

#### 2022

Francisco Gindre  
Alejandro Martín Aguirre  
Sergio Hernán  
Valenzuela Cámara

#### TIAE

#### 2022

Nelba Quintana  
Lucas Romano

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

## Tesis completa:

<http://sedici.unlp.edu.ar/handle/10915/135556>



## Tesista

María Roxana Martínez

## Directora

Dra. Rocío Rodríguez

## CoDirectora

Dra. Claudia Pons

## Asesor

Dr. Pablo Vera

## 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 principios 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.

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

## Tesis completa:

<http://sedici.unlp.edu.ar/handle/10915/134522>



## Tesista

Ariel Pasini

## Directora

Dra. Elsa Estevez

## CoDirectora

Lic. Patricia Pesado

## 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.

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

## Tesis completa:

<http://sedici.unlp.edu.ar/handle/10915/135961>



## Tesista

Jonathan Martin

## Directora

Dr. Diego Torres

## CoDirectora

Dr. Alejandro Fernández

## 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.

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

## Tesis completa:

<http://sedici.unlp.edu.ar/handle/10915/135846>



## Tesista

María José Basgall

## Directora

Marcelo Naiouf

## CoDirectora

Alberto Fernández Hilario

Tesis en cotutela, también aprobada en la Universidad de Granada (España)

## 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 sola-

pamiento 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.

# Modelado de Sistemas Colaborativos

## Tesis completa:

<http://sedici.unlp.edu.ar/handle/10915/138394>



## Tesista

Luis Mariano Bibbo

## Directoras

Claudia Pons, Roxana Giandini

## CoDirector

Gustavo H. Rossi

## 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 verifiquen 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.

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

## Tesis completa:

<http://sedici.unlp.edu.ar/handle/10915/136870>



## Tesista

Giovanni Daián Rottoli

## Directoras

Mg. Rodolfo Bertone

## Asesor

Dr. Hernán Merlino

## 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.

# Generalización del Modelado de Caden- cias de Tecleo con Contextos Finitos para su Utili- zación en Ataques de Presentación y Canal Lateral

## Tesis completa:

<http://sedici.unlp.edu.ar/handle/10915/140289>



## Tesista

Nahuel González

## Director

Dr. Jorge Salvador Ierache

## CoDirector

Dr. Waldo Hasperué

## Asesor

Dr. Enrique P. Calot

## 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.

# Cálculo Científico Distribuido sobre Clientes Móviles Indeterminados en redes MANETs

## Tesis completa:

<http://sedici.unlp.edu.ar/handle/10915/139262>



## Tesista

Pablo José Iuliano

## Directoras

Ing. Luis Marrone

## Asesor

Dr. Fernando G. Tinetti

## 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.

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

## Tesis completa:

<http://sedici.unlp.edu.ar/handle/10915/139509>



## Tesista

Marisa Daniela Panizzi

## Directora

Marcela Genero Bocco

## Codirector

Rodolfo Alfredo Bertone

## 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, activida-

des 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.

- 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.

# Proceso de Requisitos Validado Empíricamente



## **Tesista**

Gladys N. Kaplan

## **Director**

Jorge H. Doorn

## **MOTIVACIÓN**

La construcción del Proceso de Requisitos basado en Escenarios, tema central de la presente tesis, ha comenzado hace más de dos décadas. Durante este tiempo, se lo ha revisado en diferentes proyectos de investigación, mientras que simultáneamente se lo ha aplicado en numerosos casos.

La detección de algunos problemas ha generado una alerta en la calidad y consistencia de los modelos construidos y, por consiguiente, de los requisitos del software obtenidos. Su estrategia de construcción consta de tres etapas: Comprender el Universo de Discurso actual, Planificar el Universo de Discurso futuro y Explicitar los Requisitos del Software. En la secuencia de ejecución del proceso se utilizan básicamente dos modelos: el Léxico Extendido del Lenguaje (LEL) y los Escenarios, cada uno con sus particularidades. Tanto el proceso como los modelos han sido analizados empíricamente en esta tesis. Parte de los resultados obtenidos consisten en reemplazar la Actividad Derivar Escenarios Actuales con un nuevo mecanismo cognitivo que ayuda a construir una primera versión de los Escenarios de mejor calidad, ya que la actividad existente genera Escenarios con problemas de consistencia y completitud. El principal origen de estos problemas ha sido ignorar que el LEL es declarativo y los Escenarios son procedurales. La nueva heurística utiliza un mecanismo procedural e incremental por proximidad de situaciones, y, además, se nutre de todo el conocimiento existente en el macrosistema. En forma análoga, se ha detectado que la Heurística de Construcción del LEL, la cual es conducida por una lista inicial de símbolos, dificulta la detección de nuevos símbolos, afectando la completitud del glosario. En este caso también se propone reemplazarla por una nueva heurística que utiliza la lista inicial a modo recordatorio y propone identificar los símbolos por proximidad, arrojando mejoras significativas en los LEL construidos. Otros cambios son transversales, como es la incorporación de las jerarquías conceptuales y de los puntos de vista del contexto ya que afectan tanto al LEL como a los Escenarios. De esta manera se incluye información sensible para los requisitos, incrementando el nivel de detalle y la precisión de todas las representaciones. Con el objetivo de reducir los errores involuntarios, durante la descripción del LEL

y de los Escenarios, se propone una Vista Clasificación, que puede activarse por demanda, con información adicional de cada símbolo del LEL. Las restantes modificaciones son nuevos modelos agregados al proceso, el primero de ellos es la Construcción del LEL de Requisitos, lo que ha puesto de manifiesto una importante omisión en el proceso existente al no contemplar la evolución que sufre el léxico durante el proceso de Ingeniería de Requisitos. La utilización del LEL en los Escenarios futuros y en el documento de especificación de requisitos de software, paradójicamente, es una nueva fuente de ambigüedades ya que el léxico de los clientes y usuarios no es apto para describir el proceso del negocio futuro con el sistema de software en ejecución. Finalmente, el segundo agregado está relacionado con aquella información que aparece espontáneamente, pero que no tiene cabida en el modelo que se está construyendo. Esta información extemporánea requiere ser resguardada en el momento que aparece para recuperarla oportunamente, asegurando su comprensión cuando llegue el momento de incorporarla a un modelo. A tal efecto, se describe un mecanismo que ayuda el tratamiento efectivo de este tipo de información en cualquier proceso de requisitos. Todos los cambios y agregados al Proceso de Requisitos basado en Escenarios contribuyen a obtener una Especificación de Requisitos de Software de la mayor calidad posible, proporcionando modelos más completos y consistentes.

## APORTES DE LA TESIS

El aporte de la presente tesis se puede sintetizar en tres puntos. En primer lugar, el mismo Proceso de Requisitos basado en Escenarios en el cual la tesista ha tenido diferentes tipos de intervención durante más de dos décadas (ver Parte 2). En segundo lugar, el estudio empírico realizado al proceso de requisitos en general y en particular a los modelos LEL y Escenarios (ver Parte 3). Los problemas centrales detectados están relacionados con la completitud y la consistencia. Finalmente, se han propuesto un conjunto de agregados y mejoras que aumentan significativamente la calidad de los modelos y de todo el proceso (ver Parte 4).

## LÍNEAS DE I/D FUTURAS

Se enuncia a continuación una serie de temas tratados en la presente tesis que quedan abiertos para futuras investigaciones:

- Elicitación: mejorar la detección de los usuarios indispensables para garantizar que la IR cuente con toda la información necesaria. Se ha trabajado en la generación de nodos cuya distancia es el conocimiento de cada usuario. Esta información se obtiene a partir de cuestionarios. Este mecanismo no se ha probado aún.
- Validación de Escenarios: es necesario mejorar la validación de los escenarios, principalmente de los EF. Se está trabajando en agrupamientos de escenarios según su destino. Este es un aspecto sobre el que se tiene una gran confirmación experimental, el agrupamiento de los escenarios favorece la legibilidad del conjunto, sin embargo, no existe un sólo agrupamiento posible, por ejemplo, un agrupamiento pensado para la validación no es el mejor para facilitar el diseño del sistema. Se considera que la validación de los EA es muy diferente a la de los EF ya que los EA representan un contexto existente, mientras que los EF representan un contexto que será real en un tiempo. En este sentido, la validación tiene diferente nivel de complejidad según sea el usuario o los desarrolladores. Por ejemplo, los EA son más fáciles de comprender para los usuarios ya que es el contexto existente, el cual conocen. En este caso la complejidad es para los ingenieros/as de requisitos. En cambio, en el caso de los EF, la complejidad mayor es para los usuarios ya que deben imaginar el contexto futuro que aún es inexistente. Se espera probar en casos reales el uso de Puntos de Vista en el LEL y Escenarios, como, así como también las nuevas heurísticas de la Construcción del LEL y de la Derivación de EA.

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

## Tesis completa:

<http://sedici.unlp.edu.ar/handle/10915/137845>



## Tesista

Francisco Gindre

## Director

Dr. Matías Urbieta

## MOTIVACIÓN

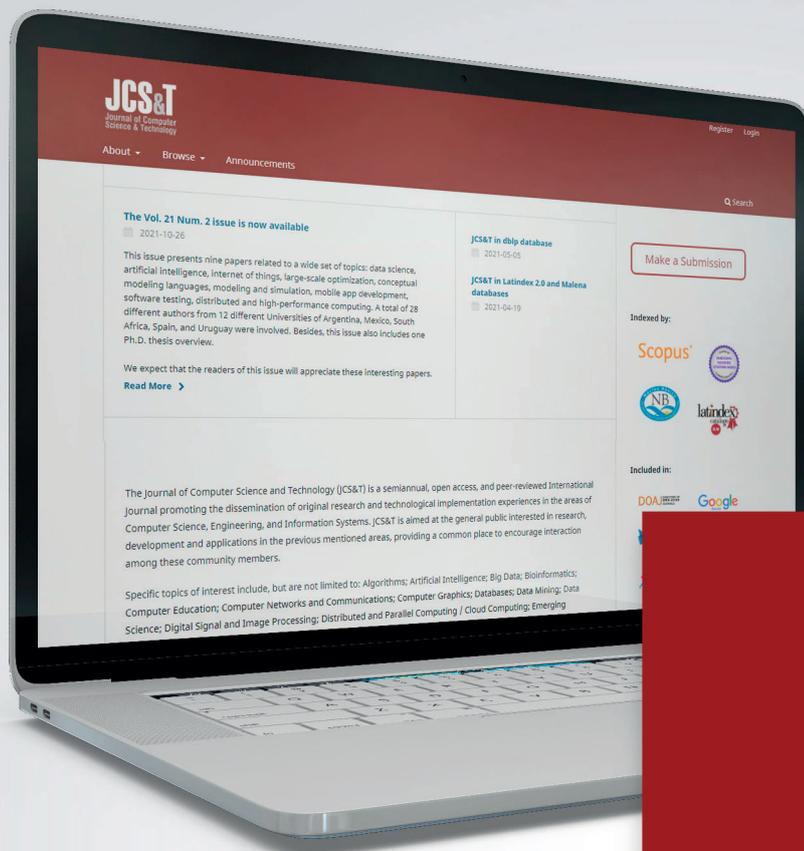
Desde su surgimiento en 2009, el uso de criptomonedas ha estado en constante crecimiento en términos de cuota de mercado y adopción. Este "boom" está a la vista de todos, y al igual que la gran mayoría de las transacciones del mundo de las finanzas descentralizadas, es de completo estado público. Pese 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



# JCS&T

## Journal of Computer Science & Technology

Vol. 21 | No. 2 | October 2021

Postgraduate Office of the School of Computer Science - UNLP

RedUNCI  
(Argentinian Universities Network  
with Computer Science Degree)

ISTEC  
(Iberoamerican Science & Technology  
Education Consortium)

ISSN  
1566-8048 (print)  
1566-6038 (online)

e-mail: [journal@lidi.info.unlp.edu.ar](mailto:journal@lidi.info.unlp.edu.ar) | Web: <http://journal.info.unlp.edu.ar>

JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY

<https://journal.info.unlp.edu.ar/>

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



## Tesista

Alejandro Martín Aguirre

## Director

Dr. Matías Urbieta

## 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.

Los objetivos específicos son:

- 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 re-

servas, 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.



2021

# TESIS Y TESISISTAS

DOCTORADO · MAESTRÍAS · ESPECIALIZACIONES



## Postgrados Acreditados en Informática en Argentina



TESIS Y TESISISTAS

POSTGRADOS ACREDITADOS  
EN INFORMÁTICA



# POSTGRADO

FACULTAD DE INFORMÁTICA

ENCONTRÁ NUESTRAS  
ÚLTIMAS PUBLICACIONES EN  
<https://postgrado.info.unlp.edu.ar/>

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

## **Tesis completa:**

<http://sedici.unlp.edu.ar/handle/10915/135839>



## **Tesista**

Nelba Quintana

## **Director**

Alejandro Gonzales

## **Asesora**

Alcira Vallejo

## **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.



inscripción abierta



# cursos septiembre

Aspectos avanzados de Ingeniería de requerimientos usando léxico extendido del lenguaje.

Metodología de la Investigación.

Introducción a la Gestión de Ciudades Inteligentes.

Minería de Textos.

Modelos y Simulación de altas prestaciones en HPC.

Wikipedia en Educación.

Sistemas de Producción Inteligente.

Tolerancia a Fallas en Sistemas de Cómputo de Altas Prestaciones.

Realidad Aumentada.

P  
O  
S  
T  
G  
R  
A  
D  
O

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

## Tesis completa:

<http://sedici.unlp.edu.ar/handle/10915/139770>



## Tesista

Sergio Hernán

Valenzuela Cámara

## Director

Dr. Alejandro Fernández

## CoDirector

Mg. Diego A. Aracena Pizarro

## 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 enfermedades 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.

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



## **Tesista**

Lucas Romano

## **Directora**

Dra. Cecilia Sanz

## **CoDirectora**

Esp. Gladys Gorga

## **MOTIVACIÓN**

La Realidad Aumentada es una tecnología emergente que permite interactuar con elementos virtuales que se agregan sobre entornos reales. En la medida en que avanza la tecnología, se facilita la incorporación de software y hardware que utiliza Realidad Aumentada en los diversos aspectos y facetas de la vida de las personas, incluso en los dispositivos móviles personales. El aplicar tecnologías de Realidad Aumentada a los escenarios educativos posibilita a los alumnos el desarrollo de habilidades tales como la capacidad espacial y mejoras en las prácticas con temas abstractos, que en el mundo real son difíciles de implementar. También se verifica la mejora de algunas variables, al utilizar estrategias pedagógicas con materiales con Realidad Aumentada, tales como: motivación, interés, atención, satisfacción, interacción, sensación de control, entre otras.

Este trabajo brinda un marco teórico de investigación acerca de la Realidad Aumentada en diferentes contextos educativos y su vinculación con el rendimiento académico, a través de variables de estudio asociadas a éste. Para ello, se realiza un estado del arte acerca del concepto de Realidad Aumentada, y de las posibilidades que ofrece cuando se la aplica a diferentes niveles educativos bajo una propuesta didáctica adecuada.

Luego, se lleva a cabo un estudio acerca del concepto de rendimiento académico y sus variables asociadas que provocan impacto en el desempeño de los alumnos. Se realiza un análisis de investigaciones que, en mayor medida, concluyen que el rendimiento académico de los alumnos mejora al utilizar tecnologías de Realidad Aumentada.

Posteriormente, se desarrolla un material educativo y se diseña un estudio de caso que utiliza tecnologías de Realidad Aumentada en el nivel universitario. Se realiza el estudio de caso y se obtienen resultados concretos en base al análisis de los datos obtenidos, por medio de diferentes instrumentos de recolección de datos implementados.

Finalmente, tanto a través de la investigación teórica como de los resultados del estudio de caso realizado, este trabajo permite concluir que es posible obtener un mejoramiento en el rendimiento académico de los alumnos, en especial en las variables motivación, interés, atención,

satisfacción y calificación, al aplicar tecnologías de Realidad Aumentada durante los procesos de enseñanza y aprendizaje.

## APORTES DE LA TESIS

La propuesta didáctica presentada, mediada a través de tecnologías con RA, se considera de gran utilidad tanto para los alumnos como para los docentes. El tener un “asistente virtual de cátedra” al cual un alumno puede acudir, de manera sencilla, a través de su dispositivo móvil se presenta como un beneficio innovador para los procesos de enseñanza y aprendizaje.

La ayuda del asistente virtual (un robot 3D animado en este caso), en el material creado con tecnologías con RA, para la resolución de ejercicios es considerado por los docentes como una herramienta valiosa y eficaz.

Se observa que el alumno genera mayor autonomía para la realización de los ejercicios, y al mismo tiempo fortalece el cambio del rol docente durante los encuentros sincrónicos, que guía y acompaña la participación de los alumnos que vienen con trabajo previo realizado. Esta propuesta puede ser provechosa especialmente en cátedras donde la matrícula de alumnos es alta, y el equipo docente escaso.

## LÍNEAS DE I/D FUTURAS

Como línea de trabajo a futuro, se considera de interés continuar con la investigación de herramientas de autor emergentes que permitan optimizar la creación de materiales aumentados con fines educativos. De esta manera, los docentes o expertos en contenidos contarán con más opciones de incorporación de contenido virtual, y funcionalidades para las diferentes plataformas de los dispositivos que los alumnos utilicen.

Otra futura línea de investigación de interés consiste en profundizar en las variables asociadas al rendimiento académico y su vinculación con las tecnologías de RA. Esta investigación obtiene resultados similares a los encontrados en la revisión sistemática de la literatura, sin embargo, la muestra no permite ser concluyente. Se aporta a sostener los resultados previos encontrados en otros estudios. Al mismo tiempo, se abre la puerta para investigar en variables de entrada que también pueden incidir en los resultados tales como género, edad y conocimientos previos.

Por último, se espera avanzar en el diseño y desarrollo de materiales educativos con tecnologías con RA en otras cátedras en el ámbito universitario, con el uso de un “asistente virtual”, que acompañe a los alumnos a realizar ejercicios, a fines de obtener resultados con una muestra más amplia y diversa de alumnos de nivel universitario.