



# Estudiantes DCC



En esta sección de la Revista estudiantes recientemente graduados del Departamento de Ciencias de la Computación (Universidad de Chile) nos cuentan, junto a sus profesores guías, sobre sus trabajos de memoria y/o tesis.



Tesis de doctorado

## Spatio-temporal Textual Data Modeling

**Estudiante:** Juglar Díaz

**Profesores guías:** Bárbara Poblete y Felipe Bravo Márquez



Me gradué de Ciencia de la Computación en la Universidad de Oriente en Santiago, Cuba. Después de graduarme estuve unos años vinculado a un centro de investigación asociado a la universidad donde investigaba en temas de minería de textos, y desarrollaba soluciones que luego eran implementadas en aplicaciones prácticas. Al mismo tiempo trabajaba como docente en materias de inteligencia artificial en la misma universidad. Mi primer contacto con el DCC de la Universidad de Chile fue a través de compañeros de pregrado que fueron aceptados en el programa de doctorado. Lo que me motivó finalmente a postular, fue el trabajo de Bárbara Poblete en temas de minería de datos en redes sociales. Bárbara Poblete y Felipe Bravo Márquez, que se incorporó luego al DCC, terminaron siendo mis profesores guías.

Mi tesis de doctorado "Spatio-temporal Textual Data Modeling" se enfocó en la representación conjunta de texto, tiempo, y espacio. La popularidad y el uso de redes sociales en dispositivos móviles con GPS, proporciona una fuente de textos enriquecidos con contexto espacio-temporal. Otros dominios, como consultas a motores de búsqueda y descripciones de incidentes delictivos, son también fuentes de textos para los que se conoce cuándo y dónde fueron generados. Texto, tiempo y espacio tienen diferentes formas de representación, por lo que no es trivial desarrollar un modelo que los represente de forma conjunta. La representación conjunta de texto, tiempo y espacio se ha basado en técnicas que ignoran la estructura secuencial de los textos además de propiedades de tiempo y espacio como

vecindad y jerarquía. Esto puede limitar la expresividad de un modelo para representar ciertos patrones. En la tesis presentamos dos nuevos modelos basados en redes neuronales para recuperación de información multimodal y modelado de lenguaje condicionado espacio-temporalmente. Los modelos propuestos presentan aplicaciones prácticas en la recuperación en texto-espacio-tiempo y en la caracterización de zonas urbanas.

Durante el doctorado, el método de trabajo fue encuentros semanales donde actualizábamos cómo íbamos en la investigación o la publicación que estuviéramos trabajando en ese momento y decidíamos cuáles eran los próximos pasos a seguir. Trabajar con Bárbara y Felipe fue una experiencia muy buena y de mucho aprendizaje constante sobre cómo se desarrolla un proyecto de investigación, cómo se escribe un artículo científico, y cómo se muestran los resultados a la comunidad científica. En el caso de la investigación en minería de datos, existe mucho entusiasmo por los nuevos avances en técnicas de aprendizaje automático basadas en redes neuronales, lo que puede guiar a un estudiante de doctorado hacia una filosofía enfocada en modelos. Si tuviera que resumir el mayor aprendizaje de mi investigación durante el doctorado, diría que es enfocarse en los datos y permitir que los datos y la información que descubramos en los datos sea lo que guíe el trabajo de investigación. Esta filosofía la llevo conmigo en mi nueva etapa profesional donde me desempeño como Científico de Datos Principal en la división MACH del Banco de Crédito e Inversiones (BCI).



Tesis de doctorado

## Engineering Interaction Requirements of Mobile Applications that Support People-Driven Collaborative Processes

**Estudiante:** Maximiliano Canché

**Profesores guías:** Sergio Ochoa y Daniel Perovich



Estudié Licenciatura en Ciencias de la Computación en la Universidad Autónoma de Yucatán, México. Entre el último par de años de mis estudios y un tiempo después de egresar, trabajé en la industria y la administración de un centro de cómputo para después realizar un Magíster en Ciencias en Tecnología Informática en el Instituto Tecnológico y de Estudios Superiores de Monterrey, México. Después de ello trabajé en la docencia en mi universidad de formación y en 2016, por interés en la investigación, decidí iniciar el doctorado en el Departamento de Ciencias de la Computación de la Universidad de Chile.

Mi tesis de doctorado, titulada “Engineering Interaction Requirements of Mobile Applications that Support People-Driven Collaborative Processes”, se enmarca en el área de ingeniería de requerimientos de sistemas colaborativos. Dicho trabajo está centrado en la identificación y validación de los requerimientos de interacciones para el diseño de aplicaciones que apoyan procesos colaborativos dirigidos por personas (PDCP, por sus siglas en inglés), cuyo flujo de trabajo no está definido al momento de su diseño ya que dependen en la práctica de variables del contexto (tales como una lista de prioridades de las tareas, disponibilidad de las personas, etc.). Un estudio inicial realizado en este trabajo, muestra que los lenguajes y notaciones usados para crear modelos visuales de los escenarios de interacción de procesos colaborativos dirigidos por personas, resultan limitados para abordar los desafíos involucrados en esta actividad de modelado. En consecuencia, proponemos un lenguaje denominado CIMoL (Computer-Mediated Interactions Modeling Language) para modelar escenarios de interacción mediados por computadora involucrados en PDCPs. Este lenguaje, junto con una herramienta de generación automática de prototipos (denominada CIMoL Modeler), permite a los ingenieros de software y *stakeholders* identificar y validar los servicios de interacción (e.g. audio y videoconferencia, envío de mensajes de texto, pizarrón electrónico, etc.) que requieren los

participantes en un PDCP con el fin de lograr un entendimiento compartido sobre el alcance del sistema en desarrollo.

Uno de los principales desafíos a los cuales me enfrenté al iniciar el doctorado, fue volver a adquirir el ritmo de estudiante de pregrado. Entre las clases, tareas, proyectos, exámenes, desarrollo del tema de tesis desde el inicio de los estudios, me surgieron complicaciones, principalmente en los dos primeros semestres del doctorado. Afortunadamente, el adaptarme de nuevo a tal ritmo me facilitó superar otros desafíos implícitos en los estudios doctorales.

El trabajo con mis profesores guías, Sergio Ochoa y Daniel Perovich fue excelente, las capacidades de ambos se evidencian en la dirección de todos los productos generados en este trabajo de tesis. Primero inicié con Sergio, cuyas ideas, disposición y conocimientos, me ayudaron a definir mi tema de investigación. Posteriormente, Daniel nos apoyó para mejorar la propuesta inicial hasta ser lo que ahora finalizamos. El camino no fue fácil, sin embargo, el apoyo de ambos fue determinante para poder lograr terminar el doctorado exitosamente. En ese camino tuve la oportunidad de crear amistad con gente del Departamento, además de vínculos con investigadores de Chile y de otros países, asistir a diversas conferencias internacionales, publicar varios artículos de investigación derivados del trabajo de tesis, asistir a una escuela de verano y realizar una pasantía de investigación.

Actualmente me encuentro trabajando en México en mi universidad de origen, la Universidad Autónoma de Yucatán, como académico en la carrera de Ingeniería de Software. Adicional a la docencia, realizo actividades de tutoría, gestión, e investigación, lo cual incluye dirección de tesis de estudiantes de Licenciatura en la que aplico muchos de los conocimientos y habilidades adquiridos durante los estudios doctorales.



Tesis de doctorado

## Socio-Technical Analysis of The Robot Operating System Ecosystem for Fostering Participation in Knowledge Sharing Activities



**Estudiante:** Pablo Estefó

**Profesores guías:** Jocelyn Simmonds, Romain Robbes y Johan Fabry

En mis últimos años de pregrado en el DCC comencé a adentrarme en el mundo de la investigación. Junto con el profesor Alexandre Bergel, trabajamos en una herramienta visual para la refactorización de tests unitarios. Este trabajo fue seleccionado en la Student Research Competition enmarcada en la Conferencia Internacional de Ingeniería de Software (ICSE), la más importante del área donde obtuve el primer lugar en la categoría pregrado. Este hito me llevó a realizar una pasantía de investigación en la École des Mines de Douai, al norte de Francia. Allí tuve mi primer acercamiento a la robótica. A mi retorno de la pasantía, recibí la invitación del profesor Johan Fabry para realizar el doctorado en el tema de Ingeniería de Software para Robótica.

Mi trabajo de doctorado se titula "Socio-Technical Analysis of The Robot Operating System Ecosystem for Fostering Participation in Knowledge Sharing Activities" (en español, "Análisis Sociotécnico del Ecosistema de Robot Operating System para fomentar la participación en actividades de intercambio de conocimientos"). El desarrollo de software para robots es una tarea compleja con desafíos particulares que la diferencian de, por ejemplo, el desarrollo de software para aplicaciones móviles u otros contextos. Mi trabajo de tesis comprende principalmente un estudio exploratorio de la experiencia de desarrollo de los desarrolladores de software para robots. A través de un estudio de campo recolectamos, analizamos y categorizamos evidencia que nos permitió resumir en seis cuellos de botella de contribución en el ecosistema el Robot Operating System (ROS). Apoyándonos en la literatura y contrastándola con la evidencia recolectada, propusimos seis recomendaciones que abordan uno o más de estos cuellos de botella.

Finalmente, implementé un recomendador de expertos para la plataforma web de preguntas y respuestas (Q&A) que usa esa comunidad: ROS Answers. El problema clásico de estos recomendadores es encontrar el usuario más idóneo para responder una pregunta dada. En mi caso le di un pequeño giro a la pregunta estableciendo el siguiente problema: encontrar un usuario calificado para responder la pregunta, pero con un nivel de participación menor. Esto último con el objetivo de distribuir mejor la carga de trabajo y fomentar la participación de usuarios menos activos. Adapté un sistema de recomendación exist

ente que se basa fuertemente en los "tags" o palabras usadas para describir preguntas. Luego de tres pruebas pude concluir que si bien recomienda al experto, no fomenta la participación de usuarios menos activos.

La principal contribución de mi doctorado fue la evidencia de aspectos sociotécnicos del ecosistema ROS, recomendaciones basadas en la literatura y la evidencia encontrada para abordar los cuellos de botella de contribución en el ecosistema de ROS y, finalmente, un enfoque novedoso para distribuir la carga de trabajo en actividades de intercambio de conocimiento en el ecosistema de ROS.

El principal desafío que tuve al comienzo del doctorado, fue que no tenía un problema definido a tratar en el mismo. Junto con esto, el estudio de ecosistemas de software estaba recién comenzando y había muy pocos estudios de la actividad de desarrollo de aplicaciones para robótica desde una perspectiva de ingeniería de software. Esto motivó un estudio de campo para explorar y evidenciar la experiencia de los desarrolladores de software para robots.

En mi camino en el doctorado, tuve la inusual experiencia de estar bajo la tutela de tres profesores guías. El profesor Johan Fabry, quien dirigió mi primer año como estudiante de doctorado, tuvo que volver a su país natal por razones personales. Luego, trabajé guiado por el profesor Romain Robbes, quien aportó una perspectiva diferente a los estudios que estaba llevando hasta ese entonces. El profesor Robbes se trasladó a la Universidad Libre de Bolzano-Bozen en el norte de Italia. La segunda mitad del doctorado lo realicé con la profesora Jocelyn Simmonds. Trabajar con ella fue una gran escuela, aprendí mucho de metodologías de trabajo académico, así como también conté con alguien que me alentó a continuar mis estudios encontrando valor en cada aporte realizado. Finalmente puedo decir que tuve la suerte de aprender de tres académicos con vocación formativa.

Actualmente estoy trabajando en Cornershop by Uber. Mi plan es continuar desarrollándome en la industria aportando con mis habilidades desarrolladas en el doctorado.



## Tesis de magister



## Business Intelligence en la producción con un enfoque estratégico, táctico y operativo para la gestión de proyectos informáticos

**Estudiante:** Eduardo Sánchez Villegas

**Profesor guía:** Hugo Beltrán Alejos

La tesis se enfoca en la importancia de la implementación de Business Intelligence en la producción de proyectos informáticos, en particular para la gestión y control de iniciativas empresariales.

El objetivo principal es desarrollar un sistema de gestión de información para supervisar la construcción de proyectos en la organización. Como resultado, se definió y construyó un sistema de paneles de gestión que contiene varios KPIs (Key Performance Indicators) para medir y mejorar los índices de productividad. Este sistema se desarrolló mediante una gestión ágil de proyectos, realizando entregas tempranas para probar y mejorar el producto final de manera evolutiva.

Los paneles de gestión registran las horas de trabajo, el presupuesto, los recursos consumidos y los entregables de los proyectos, permitiendo un monitoreo constante. Los resultados de la evaluación muestran que el sistema proporciona visibilidad del estado de los proyectos a la dirección de la empresa y facilita la toma de decisiones para mejorar los indicadores. Durante el desarrollo del proyecto se hizo un seguimiento de las métricas para validar los objetivos. Como resultado final, se adoptaron prácticas como el registro de horas, que es esencial para cualquier indicador de gestión definido.

## Tesis de magister



## A Framework for Training and Evaluating Incremental Word Embeddings from Text Data Streams

**Estudiante:** Gabriel Iturra Bocaz

**Profesor guía:** Felipe Bravo Márquez

Los *Word Embeddings* (WE) han adquirido una relevancia indispensable en diversas tareas de Procesamiento de Lenguaje Natural (NLP) y Recuperación de Información (IR), tales como clasificación de documentos, ranking y respuesta a preguntas. Sin embargo, estos modelos presentan una limitación significativa: son entrenados en entornos estáticos, lo que significa que no pueden actualizar sus conceptos en el tiempo. Esta rigidez dificulta su capacidad para adaptarse a los patrones de lenguaje en constante evolución que surgen en fuentes como redes sociales y la Web, como nuevos hashtags o nombres de marca. Para enfrentar este desafío, se han desarrollado algoritmos de Incremental WE que permiten la actualización dinámi-

ca de las representaciones de palabras a nuevos patrones de lenguaje y flujos de datos de texto.

A pesar de los avances en los algoritmos de Incremental WE, aún enfrentamos problemas importantes. Estos algoritmos suelen encontrarse en repositorios aislados, sin una interfaz común para su utilización. Además, debido al enfoque continuo o de "streaming" en su aprendizaje, no existe una estandarización o metodología clara para evaluar la calidad de los WE creados a partir de estos algoritmos.

Frente a esta problemática, hemos desarrollado RiverText, una herramienta innovadora que busca reunir las diferentes técnicas de Incremental WE en una interfaz común y de fácil uso para la comunidad científica. Adicionalmente, RiverText propone una metodología para evaluar la calidad de los WE entrenados a partir de flujos continuos de texto, que consiste en aplicar una serie de evaluaciones periódicas utilizando tareas intrínsecas de NLP.

Los resultados de este trabajo se presentaron en la 46ª Conferencia Internacional SIGIR sobre Investigación y Desarrollo en Recuperación de Información, llevada a cabo en julio de 2023 en Taipei, Taiwán. Durante esta instancia, el trabajo fue reconocido como una contribución novedosa y aplicable, que aborda un problema real y abre la puerta a nuevas interacciones entre las disciplinas de Incremental Learning, NLP e IR.



Memoria de pregrado

## Implementación de Leapfrog Triejoin sobre Estructuras de Datos Compactas



**Estudiante:** Daniela Campos  
**Profesores guías:** Gonzalo Navarro y Diego Arroyuelo

En esta memoria se abordó el problema de resolver eficientemente consultas en bases de datos de grafos, en particular Basic Graph Patterns (BGPs), que son las consultas más frecuentes.

En las bases de datos de grafos, los nodos representan entidades y las aristas entre ellos, con labels, representan propiedades. Los datos se pueden ver como un gran conjunto de triples  $(s, p, o) = (\text{sujeto}, \text{predicado}, \text{objeto})$ , por ejemplo  $(\text{Ana}, \text{jefa-de}, \text{Bruno})$  establece una arista con label “jefa-de” que va de “Ana” a “Bruno”. Un BGP consiste en un conjunto de triples donde algunas posiciones son variables, que debe encontrarse en el grafo de todas las formas posibles. Por ejemplo,  $(\text{Ana}, \text{jefa-de}, X)$ ,  $(\text{Ana}, \text{estudió-en}, Y)$ ,  $(X, \text{estudió-en}, Y)$  encuentra subordinados  $X$  de Ana que estudiaron en su misma universidad  $Y$ . Resolver estas consultas es muy costoso, y equivale a realizar *multi-joins* (y *selects*) en bases de datos relacionales.

La *worst-case optimality (wco)* es un concepto reciente que alude a un algoritmo cuyo costo es óptimo “en el peor caso”

porque es similar al tamaño del *output* en alguna base de datos. Se ha demostrado que los clásicos planes de consultas *multi-joins* haciendo *joins* de a pares no son *wco*. A pesar de ello, pocos sistemas incorporan los algoritmos *wco* para resolver sus BGPs. La razón es que éstos requieren de mucho espacio. Por ejemplo, el Leapfrog Triejoin (LTJ), el algoritmo *wco* más popular, requiere de 6 copias de los índices para funcionar, donde cada copia es un *trie* que almacena todos los triples del grafo en distinto orden  $((s, p, o), (o, p, s), \text{etc.})$ .

En esta memoria se buscó abordar este problema mediante la implementación de los *tries* utilizando estructuras de datos compactas. Concretamente, se utilizó una representación sucinta de árboles generales que utiliza dos bits por nodo y permite navegar hacia cualquier hijo en tiempo constante. Los identificadores de los nodos se almacenaron como vectores compactos. Sobre esta representación se implementó el algoritmo LTJ, que realiza búsquedas sobre los nodos a medida que baja en el árbol.

El resultado es una representación que ocupa 4 veces menos espacio que Jena-LTJ, un índice clásico *wco* que representa los 6 *tries*. Es, además, mucho más rápido. Hay estructuras aún más compactas, como el Ring, que ocupa 30% del espacio de la nueva estructura, pero es unas 15 veces más lento en la mediana. Esta nueva estructura hace posible incorporar algoritmos *wco* en implementaciones reales que manejen bases de datos de grafos masivas. El resultado de esta memoria se incluyó en un artículo que está en evaluación en una conferencia internacional de primer nivel.

Esta memoria se realizó en el marco del Instituto Milenio Fundamentos de los Datos (IMFD).



Memoria de pregrado



## Hacia un primer modelo de implicancias de diseño para videojuegos en realidad virtual considerando estrategias de inmersión e involucramiento de jugadores

**Estudiante:** Fabiola Rivera Salgado

**Profesor guía:** Francisco J. Gutiérrez

En estos últimos años la realidad virtual (RV) se ha posicionado como una herramienta emergente en diversos dominios, transformando no sólo la forma en que interactuamos con el mundo digital, sino que también en cómo se diseñan experiencias significativas para los usuarios. En el área de videojuegos, los diseñadores han aprovechado las oportunidades de interacción que ofrece el hardware y el software para crear experiencias llamativas que transporten a los jugadores al reino virtual. Gracias a la RV, los jugadores pueden explorar entornos desde nuevas perspectivas, interactuar con personajes y objetos de forma realista y experimentar una mayor sensación de presencia e inmersión en los ambientes virtuales.

La unión de la RV y los videojuegos abre un sinfín de oportunidades, pero presenta un desafío único: la ausencia de puntos de referencia de diseño establecidos. En esta línea, aún no se tiene vasto conocimiento acerca de cómo diseñar experiencias que tengan el concepto de inmersión como principal factor a considerar. Actualmente, en la literatura se han explorado alternativas del uso de esta tecnología en los campos de la medicina y educación, pero no se han realizado estudios centrados en medir la experiencia de los usuarios en un contexto de videojuegos. Es por esta razón que se ve la necesidad de definir un marco teórico adecuado para la medición del fenómeno de experiencia de usuarios con un enfoque en inmersión para videojuegos de RV.

En este trabajo de título se exploraron estrategias de inmersión e involucramiento para videojuegos que utilizan RV como principal mecanismo de interacción con el jugador. Para ello, se inició con una búsqueda en la literatura, cuyo conocimiento fue consolidado en una primera instancia del metamodelo para experiencias de juego inmersivas en RV. En esta línea, con la finalidad de dar soporte al modelo, se realizó un estudio empírico, en el cual los participantes interactuaron con videojuegos de RV y posteriormente respondieron distintos cuestionarios para evaluar los diferentes aspectos de la experiencia. Finalmente, a partir del análisis de los resultados del estudio empírico se abstraieron implicancias de diseño, las cuales pueden ser utilizadas por diseñadores y programadores para mejorar la experiencia de usuario en términos de inmersión.