



# 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.



## Empirical foundation for memory usage analysis through software visualizations

**Estudiante:** Alison Fernández Blanco  
**Profesores guías:** Alexandre Bergel  
y Juan Pablo Sandoval



Estudié Ingeniería Informática en la Universidad Mayor de San Simón en Cochabamba, Bolivia. Durante mi carrera, conocí a Alexandre Bergel, quien se interesó en uno de mis proyectos y me propuso investigar más del área, a lo cual accedí. De esta forma continúe investigando con Alexandre. Tras graduarme, trabajé en Semantics con Juan Pablo y me interesé por crear herramientas que ayudaran a mis colegas a desarrollar aplicaciones de software. Especialmente, me intrigó lo desafiante que era lidiar con problemas relacionados con el manejo de la memoria, los cuales suelen afectar el rendimiento y causar fallas en las aplicaciones de software. Motivada por mi curiosidad, opté por realizar un doctorado, con Alexandre y Juan Pablo como tutores.

En 2018, empecé el Doctorado en Computación en la Universidad de Chile. A lo largo del tiempo, me enfrenté a distintos retos, como refinar mi tema de investigación, tomar decisiones sobre evaluaciones e implementaciones, y presentar mi trabajo realizado en conferencias o revistas. Realicé varios estudios con usuarios, lo cual además de requerir bastante tiempo, me proporcionó valiosa información cualitativa. Para cumplir con mi planificación, aprendí a anticipar contratiempos y mejoré mis habilidades de organización. También, mejore mis habilidades de comunicación al realizar varias presentaciones orales para mis tutores y compañeros del IS-CLab. En cuanto a la escritura de artículos, destaco la ayuda de Renato Cerro, quien proporcionó comentarios sobre mis manuscritos en inglés y valiosos consejos para redactar artículos atractivos para los lectores.

Mi tesis doctoral se titula “Empirical foundation for memory usage analysis through software visualizations”. En el doctorado, me enfoqué en visualizaciones interactivas enfocadas en brindar soporte durante el análisis del uso de memoria de un programa. La amplia gama de técnicas visuales, los mecanismos de interacción y la información que proporcionaban

despertó mi curiosidad sobre su efectividad en la práctica. Sin embargo, noté que había poca información sobre cómo los usuarios usaban las herramientas visuales y las necesidades de los programadores al analizar el uso de la memoria. Por lo tanto, mi tesis presenta tres contribuciones principales:

- Una revisión sistemática que destaca las principales características de las herramientas visuales de análisis de memoria, los desafíos del campo y áreas de investigación prometedoras.
- La implementación de Vismep, un prototipo visual para analizar la memoria de programas en Python, junto con un estudio con usuarios para comprender cómo los programadores utilizan y perciben Vismep.
- Un estudio observacional que presenta un catálogo de 34 preguntas que los programadores plantean al analizar el consumo de memoria, además del análisis detallado del uso de las herramientas para responder a estas preguntas y las dificultades que enfrentan los profesionales en el proceso.

En mi opinión, realizar un doctorado es reconfortante, divertido y desafiante. Al tomar esta decisión, se deben considerar varios aspectos, como el área de investigación y el tutor, entre otros. En mi caso, ya tenía áreas de investigación favoritas, como *software visualization* y *software performance*. Además, fue natural elegir a Alexandre y Juan Pablo como tutores debido a nuestra buena relación en el trabajo. Sin embargo, siempre existieron desafíos, lo cual es normal cuando se trabaja en una solución innovadora y eso es lo que lo hace interesante.

Actualmente, soy profesora part-time en la Pontificia Universidad Católica de Chile y miembro del laboratorio SEISLab donde continúo investigando en mis áreas preferidas.



## A domain-independent and language-agnostic approach for crisis event detection and understanding

**Estudiante:** Hernán Sarmiento  
**Profesores guías:** Felipe Bravo Márquez y Bárbara Poblete



Estudié Ingeniería Civil Informática en la Universidad Técnica Federico Santa María. Tras titularme, decidí ejercer mi profesión trabajando de manera independiente, combinado con labores de docencia a nivel universitario y formativo. Luego de 2 años, el 2015 decidí que era hora de retomar mis estudios por lo que me matriculé en el Magíster en Ciencias mención Computación del Departamento de Ciencias de la Computación (DCC) de la Universidad de Chile. Posteriormente en el 2018, inicié el Doctorado en Computación en el mismo departamento.

Mi tesis de doctorado “A domain-independent and language-agnostic approach for crisis event detection and understanding”, se enmarca en el área de minería de datos aplicada al análisis de redes sociales durante situaciones de crisis. La investigación se centró en adquirir mayor conocimiento sobre las técnicas que permitieran comprender emergencias al combinar distintos idiomas y tipos de eventos. En esta línea, el estudio se enfocó principalmente en detectar y caracterizar estos eventos con un enfoque independiente del dominio (o tipo de evento), así como agnóstico al lenguaje.

En la gran mayoría de los trabajos publicados hasta el desarrollo de mi investigación, predominaba el uso de mensajes escritos en inglés publicados en la plataforma X (previamente llamada Twitter). Si bien estas metodologías obtenían buenos resultados para la clasificación y caracterización de eventos de crisis, existía un vacío en proponer métodos para otros idiomas donde la cantidad de recursos existentes fuera limitada. Desde el punto de vista práctico, esto incorporaba barreras extras para el análisis en redes sociales por parte de agencias de emergencias, pues en lugares donde el inglés no es la lengua nativa no era factible la incorporación de estas tecnologías.

Más allá de los desafíos algorítmicos y computacionales existentes durante la investigación desarrollada, gran parte de los problemas estaban asociados a la limitada cantidad de datos publicados durante eventos que ocurrieron en zonas geográficas donde el idioma principal fuera distinto al inglés. Afortunadamente, contamos con el apoyo del DCC y del Instituto Milenio Fundamentos de los Datos (IMFD), los cuales proveyeron la infraestructura necesaria para recolectar y almacenar nuestros propios conjuntos de datos para los distintos estudios propuestos.

Mi primer contacto con el DCC fue con colegas que habían realizado sus estudios en esta institución. Esto me motivó a querer formar nuevas redes de colaboración y conocer distintas maneras de trabajo con relación a mi formación profesional. Cuando entré a realizar el magíster, me llamó mucho la atención los estudios que había publicado Bárbara Poblete, mi profesora durante toda mi estadía de postgrado en el DCC. Principalmente aquellos que analizaban grandes volúmenes de datos en Twitter. De esta manera, consolidé ideas que quería realizar durante el magíster las cuales se profundizaron durante el doctorado. Posteriormente, el profesor Felipe Bravo Márquez se sumó como parte de la guía de mi doctorado, lo cual me ayudó a tener una mirada más profunda del análisis de redes sociales, con un enfoque ligado a diversas técnicas del área de Procesamiento de Lenguaje Natural.

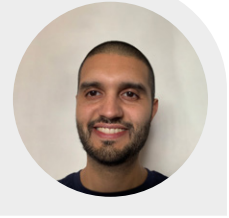
A lo largo de mi doctorado tuve la posibilidad de asistir a diferentes conferencias y workshops del área, lo cual me permitió conocer investigadores connotados, así como comprender la relevancia de sus trabajos. Sin el apoyo del DCC y del IMFD esto no habría sido posible. Más allá del apoyo para la asistencia a conferencias, el rol de estas instituciones se basó en la red investigadores y colaboradores de estas instituciones, las cuales promovieron el envío de publicaciones de excelencia que me permitieron asistir a estos eventos.

Personalmente pienso que el rol de la investigación en el desarrollo de nuestra era y cultura es fundamental. Por esta razón decidí iniciar el doctorado en el DCC, pues pude conocer la relevancia que tenían sus investigaciones a nivel mundial. Por otro lado, también siempre tuve la necesidad de conectar esas investigaciones con aplicaciones que pudieran apoyar a diversas instituciones de forma práctica. Esto gatilló que encontrara el equilibrio entre ambas labores, por lo que actualmente me desempeño como Ingeniero de Transferencia en la Dirección de Innovación y Transferencia Tecnológica del IMFD. Principalmente busco en forma constante la potencialidad de generar proyectos de investigación aplicada, por lo que tener una perspectiva y dotes de investigación me ha ayudado a plantear y planificar de mejor manera los distintos desafíos que se presentan.



## Computational models for network societies

**Estudiante:** Francisco Plana  
**Profesores guías:** Andrés Abeliuk y Jorge Pérez



Después de recibirme de ingeniero civil matemático y magíster en ciencias mención computación en la Universidad de Chile, seguí trabajando en un proyecto de investigación con mi profesor guía del magíster, Jorge Pérez, mientras en paralelo desarrollaba otros proyectos. Siempre he tenido interés en la investigación científica, la capacidad de capturar mecanismos de la modelación matemática, y otras diversas inquietudes como la música o los fenómenos sociales. La posibilidad de seguir estudios de doctorado estaba presente, y me decidí porque sentí que, sin hacer algo muy distinto de lo que ya estaba haciendo, estos estudios me darían la posibilidad de formarme como un investigador autónomo. Escogí hacerlo en el DCC porque, además de la relación que tenía con Jorge como guía, sentí que aquí tendría la libertad para realizar un proyecto ambicioso, incluso si no estaba dentro de las líneas de investigación del departamento. Y no me equivoqué.

Mi tesis se tituló "Computational models for network societies", o "Modelos computacionales para sociedades en red", y es una exploración de algunos mecanismos formales de generación de redes sociales, combinando ciencia de redes con ciencia de datos. Por un lado, estudiamos cómo la "conexión preferencial" ("preferential attachment" en inglés), un mecanismo que modela la ventaja acumulativa en el tiempo en diversos tipos de redes tanto naturales como humanas, permite estimar de forma sencilla y precisa el índice de centralidad armónica, que es un índice que puede servir para ordenar las páginas web más relevantes en una consulta. En general no es factible calcular estos índices en redes muy grandes como la Web. Por otra parte, también estudiamos las redes para compartir alimento en sociedades de cazadores-recolectores, que son muy importantes desde un punto de vista evolutivo para explicar varios atributos únicos de la especie humana, como una mayor dependencia de niños y jóvenes, cerebros grandes, dietas ricas en nutrientes, así como habilidades de forrajeo complejas, con relación a otros primates, y que tienen efectos en la organización familiar, social, política, etc. Nuestra propuesta es un modelo de optimización de redes, donde se busca minimizar tanto el riesgo de inanición individual como el colectivo, o ambos. Esto permite reproducir la organización multinivel de unas pocas familias cohesionadas observada en diversas sociedades de cazadores-recolectores, así como redes que promueven mayor o menor desigualdad de ingresos en juegos de regalos. El modelo también permite reinterpretar el origen de los patrones de reciprocidad en cazadores-recolectores como una división del trabajo dependiente del contexto social, así como aportar nuevas ideas para desarrollar modelos evolutivos más

realistas que expliquen, por ejemplo, comportamientos cooperativos que no excluyen a quienes no cooperan, como el alimento que reciben personas ancianas u otros. Las técnicas de ciencia de datos empleadas incluyeron inferencia estadística para las redes con conexión preferencial, metaheurísticas evolucionarias para computar las redes de compartir alimento óptimas, lo que involucró una serie de optimizaciones para la ejecución, además del uso de la infraestructura del Laboratorio Nacional de Computación de Alto Rendimiento (NLHPC). Elaboramos además una metodología versátil de algoritmos de agrupamiento ("clustering" en inglés) y árboles de decisión para encontrar y describir grupos de redes similares, de manera agnóstica con respecto a la forma de los grupos. Para la aplicación de estos métodos sobre diversos conjuntos de datos, desarrollamos heurísticas para fijar los parámetros de los distintos métodos.

Me parece que de las cosas más valiosas que entrega la experiencia del primer proyecto de investigación autónoma de largo plazo, dada por el doctorado, es que esta oportunidad lo empuja a uno a resolver aquellas actitudes que no te ayudan para tus objetivos. Cuando inicié el doctorado, no tenía la madurez necesaria para un proyecto ambicioso como el que me había propuesto, y tuve que ir resolviendo esto en el camino. En mi caso, esto se reforzaba por el hecho de que mi tema de investigación no es una línea de investigación habitual dentro del departamento, por lo que buena parte del trabajo fue una búsqueda personal, sin mucha retroalimentación de pares u otros profesores. Debo reconocer aquí que la confianza, compromiso y rigurosidad que mostraron mis dos supervisores, primero Jorge Pérez, y después Andrés Abeliuk, fue absolutamente clave para un desarrollo exitoso del proyecto.

Mis planes posteriores al doctorado incluyen el desarrollo de innovación de base científica en ámbitos productivos. Recientemente hemos formado nuestra empresa llamada Consystia SpA, donde producimos soluciones de I+D+i para empresas, con base en ciencia de datos y ciencia de materiales. Como comentario final, quisiera decir a aquell@s estudiantes que quieran desarrollar temas de investigación ambiciosos, poco convencionales, que quieran cambiar o impactar el mundo, por favor, ¡háganlo! el mundo necesita ideas nuevas desarrolladas con rigurosidad. Aunque no sea sencillo, si un@ es consistente en sus esfuerzos, y se apoya en sus propias fortalezas, siempre se puede salir adelante. Si son estudiantes del DCC de la Universidad de Chile, ya están en el entorno ideal para cumplir sus sueños.



## A Coq formalization of RDF and its applications



**Estudiante:** Tomás Vallejos  
**Profesores guías:** Aidan Hogan, Assia Mahboubi y Éric Tanter

RDF es un modelo de datos basado en grafos que enriquece los documentos de la World Wide Web con metadatos para que sean procesables por máquinas. Actualmente, según W3Tech, RDF es ampliamente utilizado: 39% de los sitios web usan sintaxis RDFa, mientras que el 47% usa JSON-LD; ambos formatos son sintaxis para RDF. Este nivel de adopción y la versatilidad de RDF han permitido su uso en diferentes contextos, algunos de los cuales son críticos, por ejemplo, calcular sumas de verificación o firmar digitalmente grafos RDF. Para firmar digitalmente un grafo RDF es necesario asignar una firma a cada grafo, y que tal firma sea canónica, es decir, que por un lado, les asigne la misma firma a todos los grafos con la misma estructura, y que, por el otro lado, les asigne firmas distintas a grafos con diferente estructura.

Para garantizar que un algoritmo asigna firmas canónicas a todos los grafos se puede realizar una prueba matemática. Sin embar-

go, una prueba matemática no nos libra de los bugs que puedan ocurrir a la hora de implementar el algoritmo. Que el algoritmo implementado corresponda al de la prueba matemática es crucial, en particular en contextos críticos. Para solucionar esta brecha existen herramientas de verificación formal capaces de proveer certificaciones de software sólidas, proporcionando un puente entre la programación y la lógica. En particular, el asistente de pruebas Coq ha influido en los métodos formales, la verificación de programas y matemática formal mediante el desarrollo de software confiable utilizando argumentos lógicos rigurosos.

En este trabajo presentamos CoqRDF: la primera librería mecanizada en Coq que especifica formalmente el modelo RDF, proveyendo una colección de lemas formalmente mecanizados sobre RDF. CoqRDF nos permite definir operaciones como el re-etiquetado de nodos en blanco e isomorfismo RDF y razonar sobre ellos mecánicamente. Yendo más lejos, utilizamos la librería para implementar el algoritmo  $\kappa$ -mapping, un algoritmo para asignar representaciones canónicas de grafos "RDF", y probamos en Coq que  $\kappa$ -mapping retorna la versión canónica del grafo de entrada.

Visualizamos un estándar RDF con software certificado siguiendo un enfoque mecanizado mediante el uso de asistentes de prueba. Y creemos haber contribuido con la primera piedra hacia una cadena de herramientas mecanizadas de estándares Web basada en RDF.

## Predicción de riesgo de deserción de estudiantes de primer año de Ingeniería Civil Industrial en una institución de educación superior utilizando técnicas de machine learning



**Estudiante:** Raúl Reyes  
**Profesor guía:** Benjamín Bustos

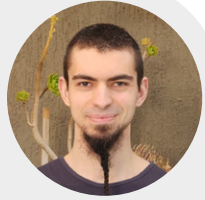
El problema de la deserción universitaria se ha convertido en un desafío prioritario tanto para las instituciones de educación superior, como para las entidades gubernamentales de Chile. Las causas de este fenómeno son múltiples y complejas, abarcando desde dificultades económicas y sociales hasta problemas vocacionales y de salud. A pesar de la implementación de diversos planes y medidas para abordar estas causas, aún no se ha logrado reducir efectivamente las tasas de deserción universitaria en el país. En este contexto, es crucial explorar enfoques innovadores y tecnológicos para abordar el problema de manera más eficiente.

Actualmente, la inteligencia artificial se aplica en una amplia variedad de campos, lo que sugiere su potencial para ser aplicado en la educación superior enfrentando problemas transversales, como la deserción universitaria. En particular, el subconjunto de la inteligencia artificial llamado aprendizaje automático (*machine learning*) puede aprovechar la enorme cantidad de datos disponibles en las instituciones educativas para identificar patrones y señales tempranas de riesgo de deserción estudiantil.

Esta tesis propone el desarrollo de una herramienta basada en técnicas de *machine learning*, para identificar a los estudiantes de primer año en riesgo de abandonar una carrera universitaria específica en una universidad estatal chilena. Al utilizar datos históricos que incluyen atributos sociodemográficos y académicos, esta herramienta no sólo permite una intervención temprana con los estudiantes en riesgo, sino que también ofrece un posible marco replicable para otras instituciones que enfrentan desafíos similares. La implementación de esta herramienta podría facilitar la detección temprana de posibles desertores y promover intervenciones oportunas para mejorar las tasas de retención universitaria.



## Predictability metrics for collaborative filtering recommender systems



**Estudiante:** Alfonso Valderrama  
**Profesor guía:** Andrés Abeliuk

Los sistemas de recomendación son fundamentales en nuestra vida digital, ayudándonos a encontrar contenido relevante, desde películas en Netflix hasta productos en Amazon. Sin embargo, entender cuán predecibles son los datos que estos sistemas procesan es un desafío poco explorado. Este estudio aborda precisamente ese problema, proponiendo nuevas métricas para medir la predictibilidad de los sistemas de recomendación basada en la complejidad estructural de los datos. En términos simples, se trata de evaluar qué tan fácil o difícil es para un algoritmo predecir las preferencias de los usuarios observando patrones en la interacción entre usuarios y artículos.

Para lograr esto, se propusieron métricas que introducen pequeñas perturbaciones en los datos para observar cómo cambia la estructura global de estos. Si la estructura exhibe pocos

cambios, significa que los datos representan interacciones que exhiben patrones predecibles, mientras que si la estructura cambia mucho con la perturbación, sugiere patrones más complejos y difíciles de predecir. Las perturbaciones a los datos fueron hechas en base a técnicas de factorización matricial, como la descomposición en valores singulares. Los resultados en datos reales de sistemas de recomendación mostraron que las métricas propuestas tienen una alta correlación con la precisión de los algoritmos de predicción, lo que valida empíricamente el enfoque propuesto.

Este trabajo es particularmente relevante, porque ofrece una herramienta que define un punto de referencia para el desempeño de los sistemas de recomendación. Esto es posible dado que las métricas de predictibilidad están definidas de manera agnóstica a los modelos algorítmicos. Además, durante la investigación se tuvo particular cuidado al definir estas métricas de forma teóricamente robusta, con el fin de estudiar la predictibilidad de forma tanto teórica como práctica.

Los resultados de este trabajo se presentaron en el workshop de Inteligencia Artificial Centrada en Datos dentro de la ACM Web Conference 2024, una conferencia internacional dedicada a discutir y presentar avances en investigación y aplicaciones relacionadas con el futuro de la web.

## Tablero digital de comunicación para niños con parálisis cerebral



**Estudiante:** Matías Bahamonde Santander  
**Profesores guías:** Francisco J. Gutiérrez

Los tableros digitales de comunicación basados en pictogramas, son uno de los Sistemas Aumentativos y/o Alternativos de Comunicación (SAAC) más utilizados por los niños con parálisis cerebral (PC). Al hacer una investigación de estos, con personal médico de esta población, se identificaron algunas problemáticas. La primera es que estas aplicaciones tienen un alto costo de entrada en algunas tiendas móviles y la segunda es que en términos de usabilidad para los niños con PC, existen insuficiencias debido a que estas implementaciones consideran a población con otro tipo de necesidades (por ejemplo, niños con trastorno del espectro autista).

Así fue que se desarrolló una aplicación en Android, que representa un tablero de comunicación basado en pictogramas, con sus funciones básicas, tales como escribir, borrar y escuchar frases, adicionando una característica de accesibilidad, dada las necesidades motrices de los usuarios, para así dar una solución a la problemática con la usabilidad. Esta se denominó tiempo de retraso de acción, el cual se define como el tiempo que el usuario debe mantener un botón dentro del tablero antes de que este se accione. Esto busca reducir respuestas no esperadas por los usuarios, dado un toque involuntario en la pantalla (efecto conocido como *midas touch*).

Este desarrollo fue evaluado mediante una prueba de concepto, por usuarios que cumplieran con el perfil de cuidador de niños con PC, en los que se incluyen al personal médico y a familiares que tuviesen bajo o relativo conocimiento de dispositivos táctiles, esto con el fin de tener un prototipo robusto, antes de ser evaluado con usuarios finales (niños con PC). Actualmente cuenta con el visto bueno de la fundación Coaniquem, para ser probado con pacientes. Por otro lado, se llevan a cabo oportunidades de mejora para este prototipo, tal como un sistema predictivo de pictogramas.