REVISTA
Bits
Edición Nº28 / Primer Semestre 2026

DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN DE LA UNIVERSIDAD DE CHILE

El nuevo marco legal de privacidad y ciberseguridad

Contenidos

dcc
CIENCIAS DE LA COMPUTACIÓN
UNIVERSIDAD DE CHILE

Revista Bits de Ciencia Nº28
ISSN 0718-8005 (versión impresa)
dcc.uchile.cl/bits
ISSN 0717-8013 (versión en línea)

Departamento de Ciencias de la Computación
Avda. Beauchef 851, 3º piso, edificio norte. Santiago, Chile. 837-0459 Santiago

dcc.uchile.cl

+56 22 9780652

revistabits@dcc.uchile.cl

/ dccuchile

Editorial

Federico Olmedo
Editor General
Revista Bits

Si vives en Chile, es muy probable que recibir llamadas *spam* se haya vuelto parte de tu día a día. De hecho, es práctica habitual que empresas de retail, call centers, bancos, etc. recurran al "mercado gris" para comprar bases de datos de información personal, originalmente recopilada con fines *muy distintos* al del telemarketing.

Esta precariedad en el manejo de la información va mucho más allá. Prueba de ello es que los incidentes tecnológicos de alto impacto hayan dejado de ser hechos aislados para volverse un patrón recurrente. En 2023, ChileCompra —el portal de compras utilizado por todas las instituciones públicas del país— quedó inoperativo durante nueve días debido a un ataque de *ransomware*. En 2021, Extranjería sufrió una "falla crítica" que resultó en la pérdida de sus registros históricos desde 1993. En 2025, un ciberataque paralizó los canales digitales del Instituto de Salud Pública (ISP), retrasando la atención ciudadana y el registro de exámenes críticos varios días.

En esta edición de Bits analizamos el nuevo marco regulatorio que busca abordar, precisamente, esta problemática y entra en vigencia plena en 2026: La Ley de Protección de Datos Personales (21.719), que devuelve al titular el control y la autonomía sobre su información personal, equiparando a Chile con estándares europeos, y la Ley Marco de Ciberseguridad (21.663), diseñada para garantizar la continuidad operativa de servicios esenciales.

Para ello invitamos a distintxs expertxs, que nos explican los puntos claves de estas normativas: Exploramos el rol de las nuevas agencias reguladoras, los desafíos prácticos que enfrentan las organizaciones para adherirse a la ley y el andamiaje técnico —desde la gobernanza de datos hasta la privacidad diferencial— necesario para implementar estos nuevos estándares de manera efectiva.

En la sección "Investigación Destacada" presentamos los resultados de dos proyectos FONDECYT: El primero, sobre algoritmos eficientes para consultas en bases de datos de grafos y, el segundo, sobre el análisis de modelos 3D de piezas de cerámica antigua. En la sección "Vinculación con el Medio" mostramos el uso de realidad virtual y videojuegos en la rehabilitación de niñxs con quemaduras, proyecto realizado en colaboración con COANIQUEM.

Cerramos la revista con la sección "Estudiantes DCC", donde egresadxs recientes comparten sus trabajos finales. Esperamos que disfruten esta edición y, como siempre, recibimos sus sugerencias en revistabits@dcc.uchile.cl. B

Estructuras de datos compactas para base de datos de grafos

Gonzalo Navarro

Doctor en Ciencias mención Computación por la Universidad de Chile. Profesor Titular del Departamento de Ciencias de la Computación de la Universidad de Chile e Investigador Asociado del Instituto Milenio Fundamentos de los Datos (IMFD) y del Centro Basal de Biotecnología y Bioingeniería (CeBiB). Líneas de investigación: diseño y análisis de algoritmos, estructuras de datos compactas, bases de datos, búsqueda en texto.

gnavarro@dcc.uchile.cl

Resumen / El proyecto Fondecyt "Compact Data Structures for Graph Databases" busca investigar algoritmos eficientes para resolver las consultas más demandantes en bases de datos de grafos, manteniendo a la vez un uso moderado de espacio. Este compromiso no es simple: los sistemas clásicos ocupan de 5 a 7.5 veces más espacio que los datos mismos, y los más recientes, que ofrecen garantías de optimalidad en la resolución de consultas, pagan su mayor eficiencia ocupando de 13 a 17 veces más espacio que los datos. Esto los hace poco atractivos en el contexto de las grandes bases de datos de grafos que están emergiendo.

Previo al proyecto habíamos demostrado que, mediante el uso de estructuras de datos compactas, era posible ofrecer una eficiencia competitiva usando mucho menos espacio, incluso prácticamente cero espacio extra. En este proyecto obtuvimos varios resultados relevantes en términos de mejorar las estructuras existentes para obtener mucho mejores tiempos y/o uso de espacio, aumentar la funcionalidad de estas estructuras para acomodar requerimientos de usos reales, y obtener nuevos resultados fundamentales en estructuras de datos compactas que impacten en el problema de implementar bases de datos de grafos.

Las *bases de datos de grafos* han suscitado gran interés con el surgimiento de grandes repositorios de información no estructurada, donde se enfatizan las relaciones entre entidades. Se han convertido en una alternativa atractiva al modelo relacional en casos donde la información no tiene estructura fija. Varios sistemas de manejo de bases de datos de grafos, prototipos, modelos y lenguajes, así como grandes repositorios como Wikidata, y sistemas propios de compañías como Apache, Neo4j, Microsoft, Oracle, Google, Facebook y otros, ilustran cuán activo es el interés en esta tecnología emergente.

Una base de datos de grafos representa la información usando un grafo (o red) con aristas (o conexiones) etiquetadas. Hay muchos modelos para representar la información de esta manera, pero en general los nodos del grafo representan objetos y las aristas entre ellos representan relaciones. Nos enfocaremos en el modelo RDF, que es uno de los más populares, donde el grafo se ve como un conjunto de *triples (s, p, o)*, donde *s* es el *sujeto* (o nodo fuente), *p* es el *predicado* (o etiqueta de la arista), y *o* es el *objeto* (o nodo destino).

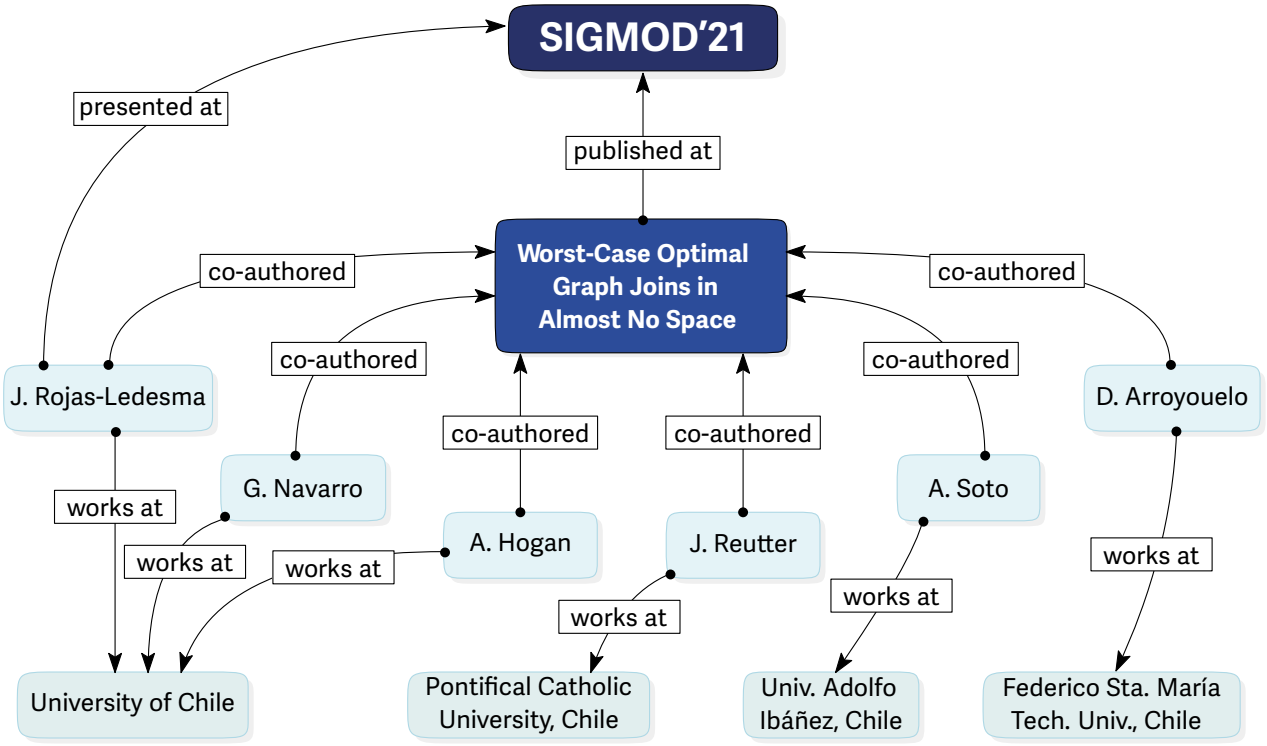


Figura 1 / Ejemplo de base de datos de grafo (incluido con permiso de su creador, Javiel Rojas-Ledesma).

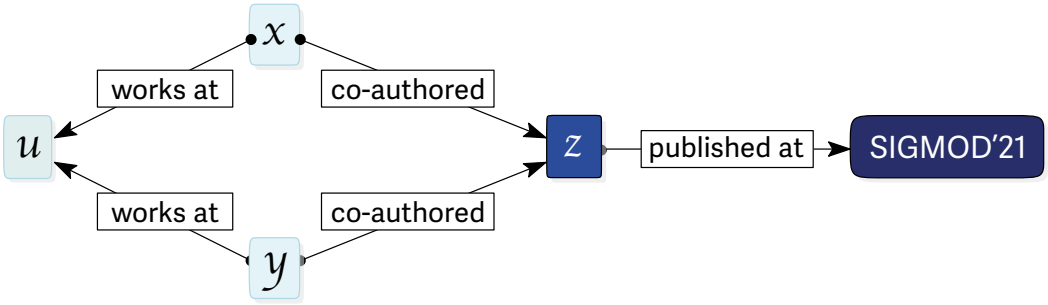


Figura 2 / Ejemplo de BGP para nuestro grafo de ejemplo (incluido con permiso de su creador, Javiel Rojas-Ledesma).

Consideremos el grafo de juguete de la Figura 1 como ejemplo. El grafo describe varios aspectos de la publicación de un artículo, como sus coautores, dónde se publicó, qué coautor lo presentó, y dónde trabajan los coautores. Por ejemplo, la arista que indica quién presentó el artículo se ve como el *triple (J. Rojas-Ledesma, presented at, SIGMOD'21)*.

Se define un *lenguaje* para consultar las bases de datos de grafos, por ejemplo un estándar muy usado es SPARQL. En casi todos, el núcleo de la consulta es un subgrafo relativamente pequeño, el cual debe ser encontrado en el grafo. En su forma más simple, el subgrafo es un único *triple pattern*, que busca una única arista en el grafo. El *triple pattern* especifica constantes o variables para el sujeto, predicado y objeto de los *triples* deseados. Cada *triple* donde calza en el grafo corresponde a *instanciar* las variables del *triple pattern*. En nuestro ejemplo, el *triple pattern (?x, co-authored, Worst...)* retorna todas las instanciaciones de *x* a coautores del artículo, es decir *x = J. Rojas-Ledesma, x = G. Navarro, x = A. Hogan*, etc.

La forma general de los subgrafos a buscar es el *Basic Graph Pattern (BGP)*, que consiste en un conjunto de *triple patterns* que comparten variables. Por cada forma en que el subgrafo resultante calce en el grafo, se deben retornar las instanciaciones correspondientes de las variables.

Considere el BGP de la Figura 2 para nuestro ejemplo, formado por 5 *triple patterns*. El BGP busca pares (*x, y*) de autores que trabajen en una misma institución *u* y hayan publicado juntos un artículo *z* en SIGMOD'21. Un resultado posible es (*x, y, u, z*) = *(J. Rojas-Ledesma, G. Navarro, University of Chile, Worst...)*.

Otro tipo de consulta que es central en SPARQL y otros lenguajes de consulta es la *regular path query (RPQ)*. Una RPQ es esencialmente una expresión regular que calza con caminos de largo variable en el grafo, de forma que la secuencia de etiquetas recorridas pertenezca al lenguaje de la expresión regular. Por ejemplo, en un grafo que represente recorridos en las estaciones del Metro de Santiago, una RPQ como *(Universidad de Chile, $L1^+ L4^+$, ?x)* me entregará todas las estaciones de metro *x* que se pueden alcanzar desde la estación Universidad de Chile usando la línea 1 (etiqueta L1) y luego la línea 4 (etiqueta L4).

Los problemas de eficiencia

La flexibilidad que ofrecen las bases de datos de grafos se paga en términos de la eficiencia en su implementación. Mientras que los *triple patterns* individuales se pueden resolver fácilmente con estructuras de datos estándar, los BGPs y RPQs son mucho más desafiantes y tienen serios problemas de eficiencia en los manejadores de bases de datos de grafos (por ejemplo, es normal que algunas consultas tomen minutos para resolverse). Es normal ver BGPs reales con hasta 20 *triple patterns*. Si bien los BGPs se pueden traducir a selecciones y *multijoins* en bases de datos relacionales (viendo el grafo como una relación de 3 columnas, (*s, p, o*)), es raro ver *joins* de tantas tablas en consultas de bases de datos relacionales, y los manejadores no son tan eficientes para este tipo de casos. Las RPQs, que no se pueden traducir directamente al álgebra relacional, son aún más costosas de resolver que los BGPs. Los lenguajes como SPARQL también permiten operaciones equivalentes a proyecciones, uniones y otras, pero el cuello de botella de la eficiencia son las BGPs y RPQs. Por eso este proyecto se centra en esos dos tipos de consultas.

Un avance muy importante en la resolución de *multijoins* fue el desarrollo de los algoritmos *worst-case optimal (wco)*. Un algoritmo de *join* es *wco* si su costo es proporcional a la llamada *cota AGM*, es decir, el máximo tamaño posible de su resultado en alguna base de datos con las mismas tablas y tamaños que las de la consulta. Se ha demostrado que las

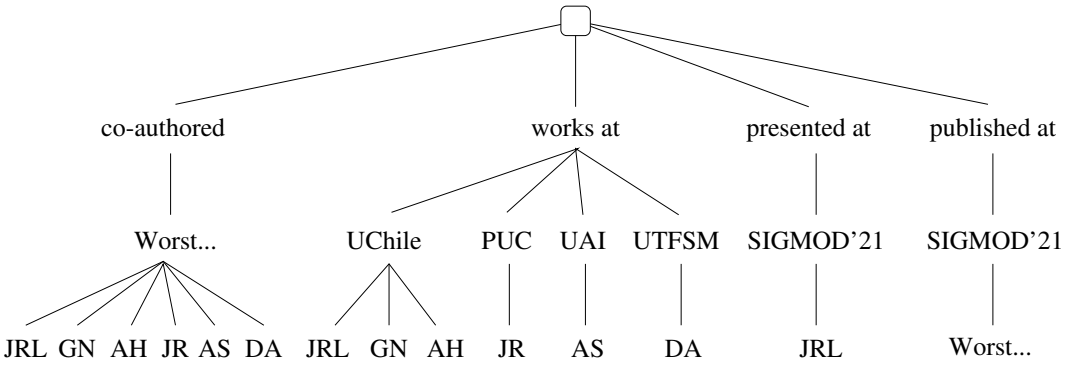


Figura 3 / El *trie* para el orden (*p, o, s*) en nuestro ejemplo (con nombres abreviados).

técnicas usadas desde la década de los sesenta por los manejadores de bases de datos relacionales, donde los *joins* se van haciendo siempre de a dos tablas, son necesariamente no-wco. Al mismo tiempo, se desarrollaron varios algoritmos *wco*, y se trasladaron las mismas técnicas a bases de datos de grafos, donde son particularmente relevantes porque los *multijoins* (o sea, los BGPs) tienden a ser grandes y complejos. Los algoritmos *wco* han demostrado ser más eficientes que los tradicionales en esos BGPs complejos, especialmente en los que contienen ciclos.

Esta mejora en tiempo tiene un costo en espacio, sin embargo. Por ejemplo, el algoritmo *wco* más popular, *Leapfrog Triejoin (LTJ)*, requiere indexar las filas de cada tabla como una secuencia de valores insertándolas en *tries*, ¡en cada posible orden de los atributos! Es decir, una tabla con d columnas debe almacenarse en $d!$ *tries*, donde $d! = d \cdot (d-1) \cdot (d-2) \cdot \ldots \cdot 1$. En el caso de bases de datos de grafos, los *triples* (*s, p, o*) se deben guardar, en los 3! = 6 órdenes posibles, en 6 *tries*, es decir, sextuplicando el espacio. La Figura 3 muestra el *trie* para el orden (*p, o, s*) en nuestro ejemplo.

Otros algoritmos *wco* tienen problemas similares de espacio. Esto es particularmente desafortunado al querer manejar grandes repositorios de datos no estructurados usando grafos, y dificulta la adopción de estrategias *wco* y rápidas para resolver consultas complejas. Como ilustración, la Wikidata contiene unos 14 mil millones de *triples*, por lo que seis copias, usando 32 bits por elemento, requieren más de un terabyte de almacenamiento.

Este es un problema donde las *Estructuras de Datos Compactas (EDCs)* pueden jugar un rol crucial. Las EDCs buscan representar tanto los datos como las estructuras que se necesiten sobre ellos en espacio cercano a la *entropía*, o cantidad de información que hay en los datos. Este es hoy un campo de investigación relativamente maduro, que ha resultado en investigación de primer nivel y aplicaciones reales, con bibliotecas de software profesionales que ofrecen representaciones compactas de varios tipos de datos, como vectores de bits, secuencias, árboles, grafos, matrices, grillas de puntos, textos, y varios otros. Las EDCs han sido muy exitosas en reducir el tamaño de varias estructuras de datos relevantes en órdenes de magnitud, así como en expandir la funcionalidad de representaciones de datos, manteniendo su espacio cercano a su entropía.

Los sistemas más recientes pagan su mayor eficiencia ocupando de 13 a 17 veces más espacio que los datos.

Es entonces muy natural aplicar EDCs al problema de implementar algoritmos *wco* en bases de datos de grafos, con el objetivo de mantener su eficiencia en tiempo mientras se elimina la redundancia. Nuestra investigación anterior a la formulación de este proyecto ya había mostrado que el uso de EDCs realmente lograba reducir el espacio que necesitan los algoritmos *wco* en bases de datos de grafos.

Este proyecto se centró en tres aspectos: (1) mejorar las soluciones compactas actuales para BGPs y RPQs; (2) extender esas soluciones para resolver otras operaciones relevantes y soportar modelos más generales; y (3) investigar en problemas básicos de EDCs que deriven en mejoras para las representaciones compactas de bases de datos de grafos. A continuación describimos los resultados más relevantes obtenidos en cada aspecto.

Mejores soluciones

El Ring. Una de las soluciones existentes antes del proyecto era el *Ring*, el cual lograba usar el espacio requerido para mantener *una* copia de los datos y con ello simular los seis *tries* al implementar el algoritmo LTJ. El Ring se había publicado en *SIGMOD'21*, y durante el proyecto se gestó la versión de revista, la que se publicó en *ACM Transactions on Database Systems* en 2024.

La razón por la cual LTJ necesita los seis *tries* es que debe ser capaz de instanciar los *triples* (*s, p, o*) en cualquier orden, por ejemplo primero el predicado, luego el objeto, y finalmente el sujeto (como en nuestro ejemplo). La idea clave del Ring es una estructura de datos que permite ver los *triples* (*s, p, o*) como una secuencia circular y bidireccional. Eso permite navegarlos en cualquier orden con un solo ring. La Figura 4 (extraída del mencionado artículo) ilustra la idea.

Durante el proyecto se extendió el Ring para permitir dinamismo, es decir, poder agregar y eliminar aristas en el grafo sin tener que reconstruir la estructura desde cero. Esto es esencial para su adopción en sistemas reales. También se aprovecharon sus estructuras compactas subyacentes para encontrar mejores órdenes de instanciación de las variables, superando notoriamente al Ring original. Estos resultados aparecerán en Information Systems, 2026. Otras extensiones se describen más adelante.

Compact LTJ. Otra innovación que se realizó durante el proyecto fue, en busca de una mayor eficiencia, realizar una implementación de los seis *tries* utilizando estructuras de datos compactas. Además de la representación compacta en sí, se estudió el problema de cuál es el mejor orden para instanciar las variables de la consulta, mostrando que es mejor ir eligiendo progresivamente la siguiente variable a medida que progresa la resolución, en vez de fijar el orden de antemano. La estructura resultante fue llamada *CompactLTJ*, o *CLTJ*. Se desarrolló una variante más compacta, que elimina parte de los seis *tries*, a cambio de a veces tener que pasar de un *trie* a otro durante la resolución de las consultas. Esta variante menor se llama *xCLTJ*. Finalmente, se agregó dinamismo a esta versión también. El CLTJ se publicó en *GRADES-NDA'24* (satélite de *SIGMOD*) y la versión de revista en *The VLDB Journal*, 2025.

La Figura 5 pone nuestros resultados en contexto. Además de nuestras mejores variantes del Ring en rojo y de CLTJ en azul, se incluyen sistemas establecidos *no wco* (Virtuoso, RDF-3X, Blazegraph), otros wco (Jena LTJ, MillenniumDB) que ocupan bastante más espacio, y otros que combinan estrategias *wco* y *no wco* (UmbraDB). Se muestran variantes del Ring y de CLTJ, que ocupan mucho menos espacio, el

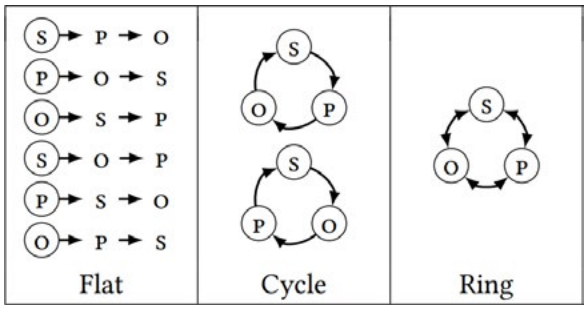


Figura 4 / El concepto del Ring: "Flat" se refiere a los seis órdenes que se deben indexar usando *tries*, "Cycle" a una estructura existente que era circular pero unidireccional (por lo cual necesita dos copias para implementar LTJ), y a la derecha el Ring.

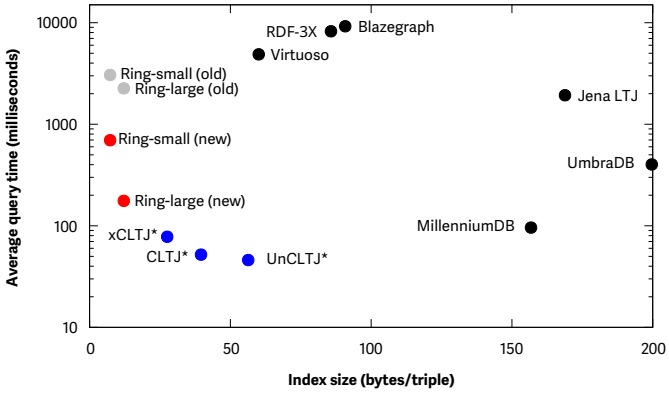


Figura 5 / Nuestros resultados en contexto, mostrando el espacio (eje *x*) y el tiempo (eje *y*) de diversos sistemas conocidos al resolver consultas de BGPs reales en un subconjunto de mil millones de nodos de Wikidata.

Ring obteniendo el mínimo y CLTJ siendo mucho más rápido y dominando a todos los demás sistemas.

RPQs. Previo al proyecto ya se habían diseñado variantes del Ring que permitían resolver RPQs eficientemente, usando la estructura para buscar caminos en el grafo. En el proyecto se mostró que esto podía lograrse de forma más eficiente con estructuras compactas simplificadas que requerían menos espacio y tiempo. Por otro lado, la estructura original, más potente, se explotó para desarrollar mejores heurísticas en la resolución de RPQs. La idea central es que, en vez de buscar los caminos de una punta a la otra, se parta de algún punto intermedio cuya etiqueta no sea muy frecuente, y a partir de sus (pocas) ocurrencias completar el camino en las dos direcciones. La estructura

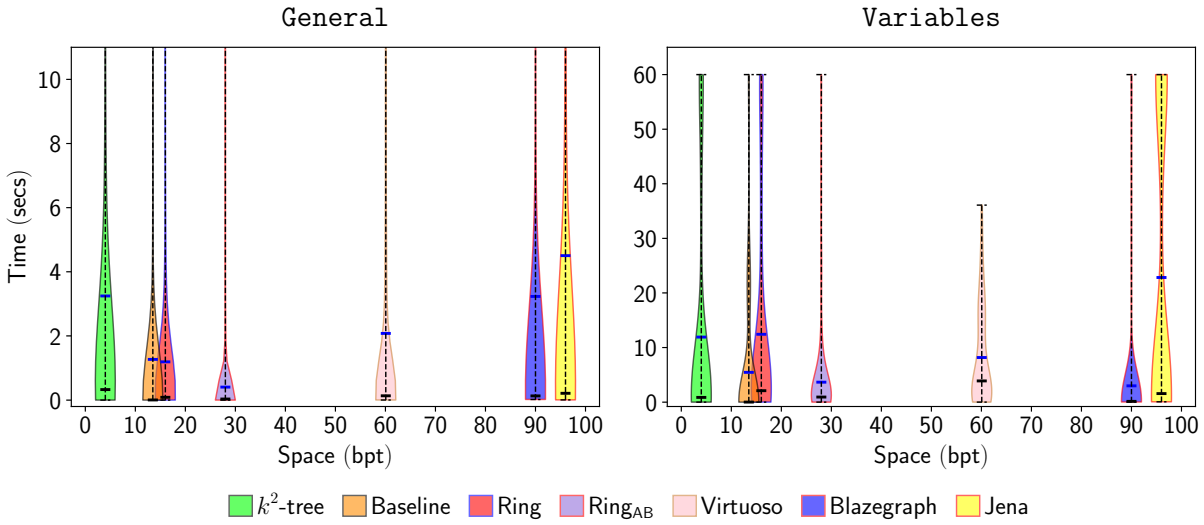


Figura 6 / Nuestros resultados sobre RPQs reales recibidas en Wikidata. A la izquierda se consideran todas las RPQs, a la derecha sólo las más difíciles (donde los caminos pueden empezar y terminar en cualquier nodo). El eje *x* es el espacio en bytes por *triple* y el eje *y* es el tiempo.

Estudiamos cuál es el mejor orden para instanciar las variables de la consulta, mostrando que es mejor ir eligiendo progresivamente la siguiente variable a medida que progresa la resolución, en vez de fijar el orden de antemano.

compacta del Ring permite estimar eficientemente un buen punto intermedio de partida, y así resolver de mucho mejor forma las RPQs. El resultado se publicó en 2024 en *The VLDB Journal*.

Por otro lado, se desarrolló una nueva línea para resolver las RPQs, que consiste en interpretar cada etiqueta distinta como una matriz booleana y dispersa, que marca de qué nodos a qué nodos del grafo se puede ir por esa etiqueta. Se mostró que las operaciones de las expresiones regulares se pueden traducir a operaciones de suma, multiplicación y clausura transitiva de matrices booleanas. Se desarrollaron nuevos algoritmos y representaciones compactas para matrices booleanas, así como optimizadores de consultas de RPQs que las traducen de la forma más promisoria a esta álgebra de matrices. Los resultados demostraron que esta representación es mucho más compacta que las anteriores y competitiva para las RPQs más difíciles. El resultado se publicó también en *The VLDB Journal*, en 2025.

La Figura 6 (extraída del mencionado artículo) muestra los resultados sobre RPQs reales recibidas en Wikidata. En cada color se ve el histograma de cuántas consultas requirieron ese tiempo. El verde es una representación compacta de matrices booleanas y el anaranjado una no tan compacta pero más rápida. El rojo es el Ring original y el lila el desarrollado en el proyecto. Los demás, usando mucho más espacio, son sistemas reales que resuelven RPQs.

Nuevas funcionalidades

El Ring en el modelo relacional. En la misma publicación de revista del Ring, se estudió cómo extenderlo al modelo relacional, donde las tablas pueden tener más de 3 atributos (como (*s, p, o*) en los grafos). En dimensiones *d* mayor a 3, se necesitan varios Rings para poder implementar LTJ. Se encontraron las cantidades de Rings que se necesitan para los primeros valores de *d*, demostrando que eran muy inferiores a la cantidad de *tries* necesarios, por ejemplo se necesitan sólo 7 Rings para $d = 6$, en vez de los $6! = 720$ *tries* que se

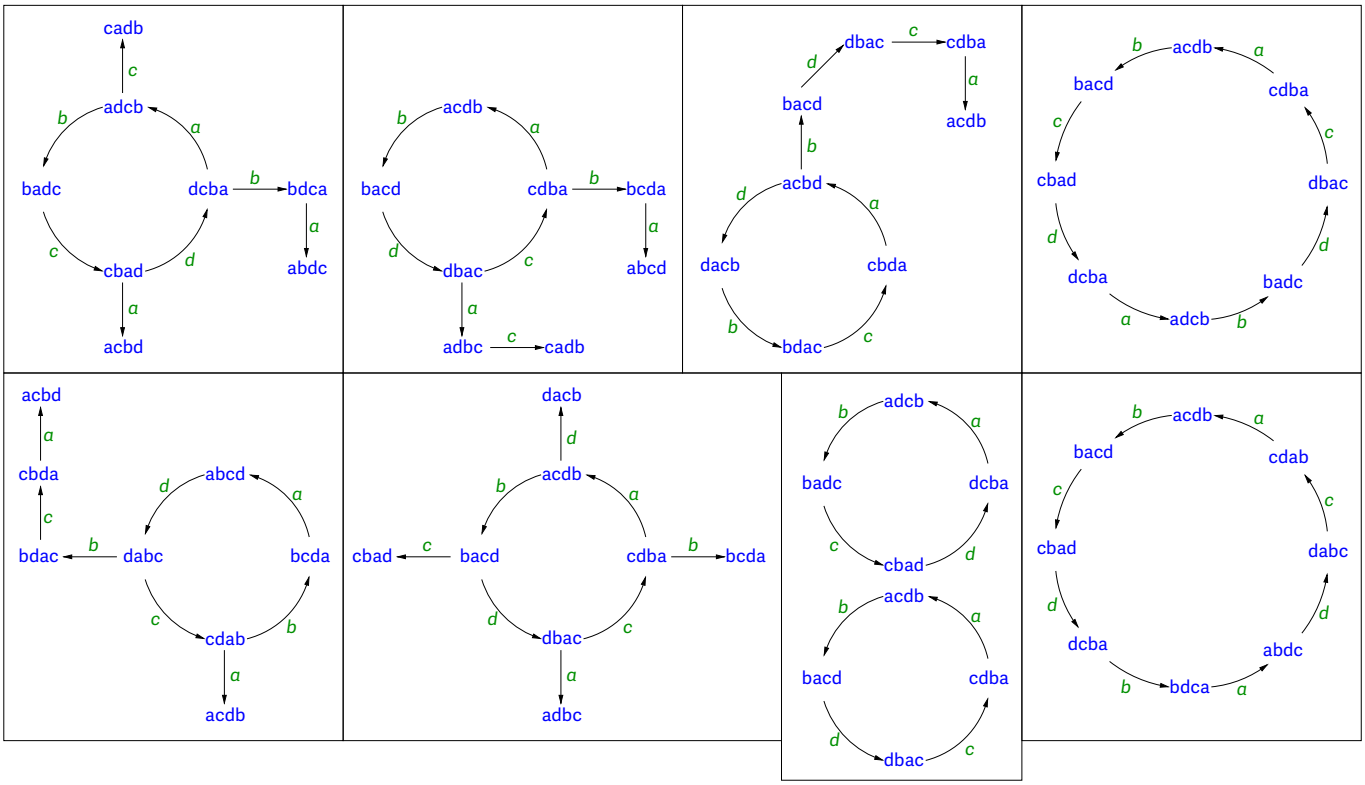


Figura 7 / Las formas de todos los order graphs de tamaño óptimo para $d = 4$, donde los atributos son $\{a, b, c, d\}$.

requerirían. Esto hace mucho más factible el uso de algoritmos *wco* en bases de datos relacionales. Encontrar la fórmula exacta para cualquier d ha quedado como problema abierto, sólo se encontraron cotas superiores e inferiores.

Un desarrollo muy intrigante fue la generalización del concepto de Ring, en dimensiones superiores, al de *order graphs*, donde los Rings forman sólo un tipo particular de *order graph* que consiste en varios ciclos de largo d. En un *order graph*, cada arista representa una columna de la tabla que se indexa en un cierto orden, obteniendo un nuevo orden de atributos. Se establecen condiciones precisas para que un *order graph* pueda implementar LTJ en una dimensión d. La Figura 7 muestra las curiosas formas de todos los *order graphs* de tamaño óptimo (8 columnas, equivalentes a dos Rings) para $d = 4$. La solución con dos Rings es la tercera de abajo.

Los order graphs resultan ser efectivamente más potentes que los Rings: para $d = 5$ se necesitan indexar 25 columnas usando Rings, mientras que los *order graphs* permiten implementar LTJ con sólo 20. Este es un muy interesante tema de investigación que queda abierto.

Grafos multimodales. Una importante extensión de las capacidades del Ring desarrollada en el proyecto es la de permitir *grafos multimodales*. Esto se refiere a que los nodos pueden tener un cierto significado (por ejemplo, representar puntos geográficos, o imágenes) y se quiere extender los BGPs con cláusulas que consulten sobre aspectos relacionados con ese significado.

Una extensión que se realizó fue permitir que los nodos pertenezcan a un *espacio métrico*, sobre el cual se pueden representar distintos tipos de cercanía entre nodos (por ejemplo, cercanía geográfica o similitud de imágenes). Se permite en las consultas que a los BGPs habituales se les agreguen cláusulas que indican que un nodo debe estar entre los más similares a otro nodo. Por ejemplo, podríamos extender nuestro BGP de ejemplo de modo que, en vez de requerir que los coautores x e y trabajen en una misma universidad u, trabajen en universidades u y v que sean geográficamente cercanas. Se mostró que el Ring puede extenderse naturalmente para implementar este modelo extendido y que, bajo ciertas condiciones, la extensión de LTJ sigue siendo *wco* y eficiente en la práctica. El resultado se publicó en *SIGMOD'24*. También se mostró cómo introducir nuevos elementos en las

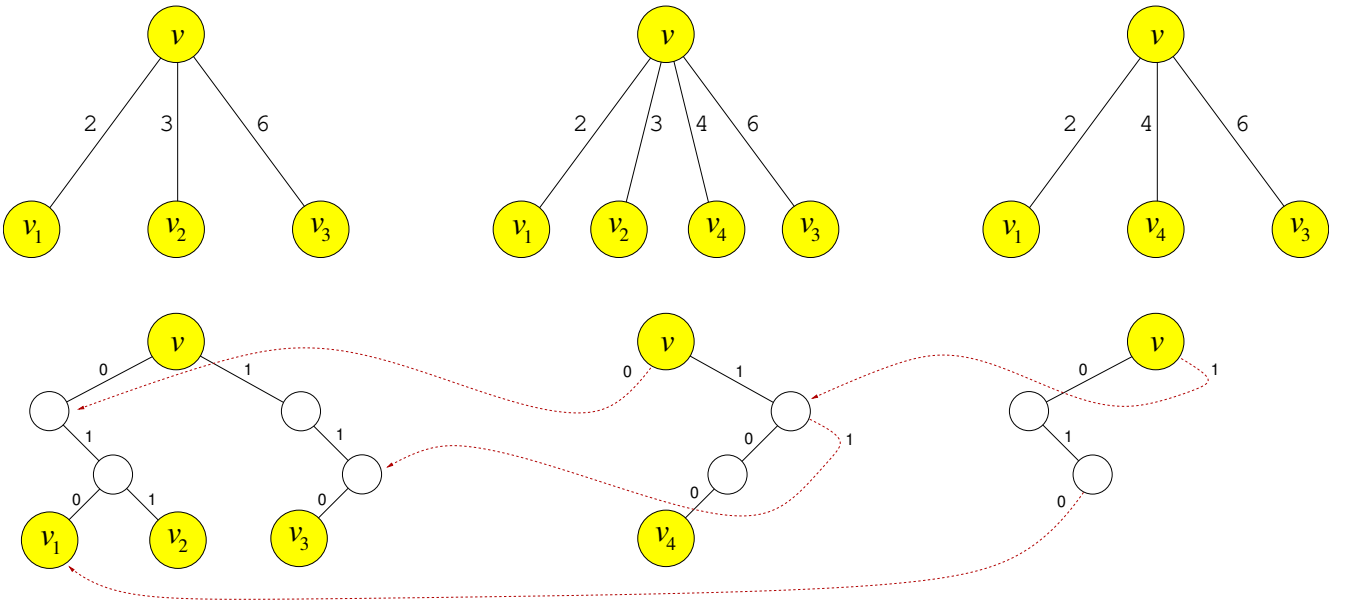


Figura 8 / Representación de los hijos de un nodo del *trie* a medida que cambian con el tiempo (primero la inserción del 4 y luego el borrado del 3), mediante un *trie* binario versionado.

consultas y no sólo referirse a objetos ya conocidos en el grafo (por ejemplo, entregar al sistema una imagen y pedir que los nodos sean parecidos a ella). Este resultado se publicó en *ISWC'25*.

Otra extensión que se estudió fue la que permite expresar que los nodos viven en un espacio topológico, donde unos pueden contener a otros, ser adyacentes, etc. Esto permite modelar, por ejemplo, subdivisiones administrativas de países. Se desarrollaron nuevas estructuras compactas para poder incorporar eficientemente a los BGPs cláusulas que indican que estos tipos de relaciones deben valer entre nodos de la consulta. Se mostró que la extensión resultante del algoritmo LTJ sigue siendo *wco*, y que el sistema resultante es eficiente en espacio y tiempo. El resultado se publicó en *WWW'25*.

Grafos temporales. Una importante extensión al modelo de grafos es permitir que las aristas tengan un intervalo temporal de validez, y que las consultas puedan poner condiciones sobre éstos. No existe a la fecha un consenso sobre cómo se puede incorporar el tiempo en las BGPs o RPQs. Se determinó concentrarse en BGPs y modelar el tiempo como una cuarta componente de los *triples*, que ahora son tuplas (s, p, o, $[t_i, t_f]$) que indican que el *triple* (s, p, o) existió entre los instantes t_i y t_f. Las consultas también tienen cuatro elementos, y pueden poner constantes o variables en el cuarto. Este esquema es muy flexible y permite expresar cosas como buscar un BGP en un instante específico, o que existiera en algún instante de un rango de tiempo, o en todo instante, o que parte del BGP exista antes que otra parte, o encontrar los cambios que tuvo el grafo entre dos instantes de tiempo (útil para sistemas versionados, por ejemplo), y en general manejar el tiempo como una componente más.

Se pueden usar estructuras de datos *persistentes* (que recuerdan su estado en cualquier momento del pasado) para implementar un LTJ extendido a grafos temporales. La Figura 8 ilustra la idea de representar los hijos de un nodo del *trie*, que van sufriendo cambios con el tiempo, mediante un *trie* binario que reutiliza partes de las versiones anteriores para evitar almacenar un nuevo *trie* para cada nuevo instante de tiempo.

Es un desafío, sin embargo, representar estas estructuras persistentes en poco espacio y además lograr tiempo *wco* al implementar LTJ, pues la forma inicial identificada para lograr ser *wco* era expandir cada tupla a muchas, cada una válida en un solo instante temporal. Se consiguió reducir ese espacio inicialmente cuadrático (lo que es totalmente impráctico) a lineal en el tamaño del grafo temporal, mediante representar esos *tries* persistentes en forma comprimida. La implementación demostró que el índice resultante es práctico en espacio y en tiempo, y que la solución ofrece mejor eficiencia que las alternativas existentes. Se ha enviado para publicación a una conferencia de primer nivel.

Esta representación [de las etiquetas de un grafo como matrices booleanas dispersas] es mucho más compacta que las anteriores, y competitiva para las regular path queries más difíciles.

Avances en EDCs

Finalmente, se obtuvieron muchos resultados a lo largo del proyecto sobre mejoras en EDCs fundamentales, las cuales impactan en las soluciones para bases de datos de grafos. Se eligen sólo tres de ellas para este artículo.

Vectores de bits dinámicos. Los vectores de bits son fundamentales en la implementación de casi todas las EDCs. Se necesita que ofrezcan unas pocas primitivas básicas, las cuales requieren tiempo constante cuando los bits no se modifican. Implementar consultas en un vector donde se puedan insertar y borrar bits, en cambio, es demostrablemente más lento, un orden de magnitud en la práctica. Esto se traslada a muchas de las estructuras compactas usadas en el proyecto. Por ejemplo, es posible mantener un Ring o un CLTJ sobre un grafo donde se insertan y borran aristas, lo cual debe permitirse en cualquier escenario real. Pero al permitir el dinamismo, las estructuras se vuelven hasta 10 veces más lentas.

Se observó que, sin embargo, la cantidad de modificaciones es muy pequeña en la práctica comparada con la cantidad de operaciones que se hacen sobre las estructuras. Con esta premisa, se diseñó una estructura de vectores de bits dinámicos cuyo desempeño se adapta naturalmente y sin intervención a la frecuencia de las inserciones y borrados. Este desempeño adaptable se demostró formal y empíricamente. Los vectores de bits se utilizaron para diseñar versiones dinámicas del Ring y de CLTJ cuyo desempeño empeora muy levemente al permitir tasas de modificaciones en el grafo incluso muy superiores a las que ocurren en aplicaciones reales, con lo cual se solucionó un problema de larga data en esta área. Los resultados se publicaron en dos revistas, además del artículo de CLTJ.

Matrices dispersas. La traducción de RPQs a operaciones en un álgebra de matrices derivó en una investigación cuyo interés va más allá de implementar RPQs, pues las matrices booleanas aparecen en muchas otras aplicaciones, por ejemplo de Aprendizaje de Máquina. Se comparó experimentalmente el desempeño de nuestras implementaciones de matrices dispersas con otros sistemas conocidos, demostrando que son competitivas y ocupan menos espacio. Se diseñaron mejores representaciones de la versión más comprimida, en términos de que las operaciones como la multiplicación sean más amigables con el *cache*, obteniendo mejoras de hasta 60% con respecto a las usadas anteriormente. A su vez, esta representación amigable con el *cache* tiene aplicaciones más allá de matrices, pues puede también representar árboles.

Recuperación de documentos en texto comprimido. Un caso muy relevante de grafos multimodales es aquel en que los nodos representan texto, y se quiere enriquecer los BGPs con búsquedas de palabras o subcadenas en el texto. Para implementar esta funcionalidad se requiere de índices para texto que puedan buscar una subcadena y listar eficientemente todos los nodos que la contienen. Este problema se llama *recuperación de documentos*. Durante el proyecto se trabajó en un nuevo índice para recuperación de los documentos donde más aparece una subcadena que ofrece un compromiso novedoso entre el espacio que ocupa y el tiempo que requiere para responder consultas. Este resultado, que será relevante a futuro para implementar este tipo de funcionalidad, se publicó en *SODA'25*. Asimismo se obtuvo una mejora importante en el espacio ocupado por un índice para textos muy compresibles, el cual reporta muy eficientemente las posiciones individuales donde ocurre una subcadena buscada, el cual también puede utilizarse para resolver este problema. Este resultado se publicó en *ACM Transactions on Algorithms* en 2025.

Conclusiones

A lo largo del proyecto se han obtenido una importante cantidad de resultados relevantes para mejorar el estado del arte en consultar bases de datos de grafos usando espacio de memoria reducido, desde mejorar los algoritmos y representaciones compactas existentes a extenderlos a nuevas funcionalidades, así como investigación en problemas básicos de estructuras compactas que impactan en las soluciones a este problema. Sólo se mencionaron las publicaciones de más alto nivel obtenidas; el total suma 12 publicaciones en revistas y 14 en conferencias, todas internacionales. Durante el proyecto se colaboró con investigadores de Chile, España, Brasil, Estados Unidos, Italia y Canadá. Tres estudiantes de doctorado, seis de magíster y dos de pregrado trabajaron (o aún trabajan) en el proyecto. Aparte de las publicaciones, los resultados se divulgaron tanto en medios escritos como las *Communications of the ACM* en 2024 y el libro "The Expanding World of Compressed Data" en 2025, como en charlas invitadas en conferencias como *ICDT'23* (Atenas), *ACDA'25* (Montreal) y el workshop *25 Years of the BWT* (Venecia, 2025). B

Algoritmos para el etiquetado y búsqueda en modelos 3D de cerámica antigua

Benjamín Bustos

Doctor en Ciencias Naturales por la Universidad de Konstanz, Alemania (2006). Profesor Titular del Departamento de Ciencias de la Computación de la Universidad de Chile e Investigador Asociado del Instituto Milenio Fundamentos de los Datos (IMFD). Líneas de investigación: recuperación de información multimedia basada en contenido y búsqueda por similitud.

bebustos@dcc.uchile.cl

Iván Sipirán

Doctor en Ciencias mención Computación por la Universidad de Chile. Profesor Asistente del Departamento de Ciencias de la Computación de la Universidad de Chile e Investigador Principal del Centro Nacional de Inteligencia Artificial (CENIA). Líneas de investigación: visión computacional 3D y computación gráfica aplicada a la herencia cultural.

isipiran@dcc.uchile.cl

Resumen / En el Proyecto Fondecyt Regular 1230448, se investigaron métodos para el análisis de modelos 3D provenientes de escaneos digitales de cerámica antigua. Este tipo de objetos, aparte de la geometría 3D, poseen diversos patrones y motivos pintados en su superficie. El análisis de estos patrones puede ayudar a entender el contexto y la semántica de la cerámica antigua. Sin embargo, realizar el análisis de estos objetos digitalizados de cerámica antigua usando el computador no es una tarea trivial, dado que pueden sufrir de diversos tipos de defectos como partes faltantes, partes astilladas, erosión en la superficie, o también incluso errores que se producen durante el proceso de digitalización. En el proyecto investigamos técnicas para asignar etiquetas textuales a los dibujos en la superficie de la cerámica, para detectar automáticamente en qué partes de la cerámica aparecen estos motivos, y para realizar búsquedas en una colección digitalizada de estos objetos.

Introducción

En arqueología digital, un problema relevante es el estudio de modelos 3D provenientes de escaneos digitales de cerámica antigua. La digitalización se puede realizar utilizando, por ejemplo, un escáner de luz estructurada, que logra obtener modelos 3D de alta calidad. El proceso de digitalización de estos objetos permite obtener lo que se conoce como "gemelos digitales", los cuales nos ayudan a estudiar estos objetos sin peligro de dañarlos o alterarlos, ya que se utiliza el computador para analizar el gemelo digital sin necesidad de manipular el objeto real. Además, la digitalización permite la preservación de estos objetos para el futuro, y su estudio nos permite comprender de mejor forma las culturas del pasado.

En el caso de la cerámica antigua, aparte del modelo 3D también se puede digitalizar la pintura en la superficie de estas cerámicas, lo que permite tener una representación en 2D y 3D de estos objetos. Si además contamos con una descripción textual del objeto, tenemos finalmente una representación multimodal del objeto compuesto por texto, imagen y objeto 3D. Dentro de las tareas relevantes que se pueden definir para el estudio de estos objetos, se encuentra, por ejemplo, la reconstrucción de partes faltantes, ya sea por pérdida, astillado, erosión o errores durante el proceso de digitalización, y también se encuentra el estudio de los motivos dibujados en las cerámicas. En la actualidad, estas tareas son realizadas en forma manual por personas expertas en el dominio, lo que es un proceso lento y tedioso, y que sólo permite analizar unos pocos objetos a la vez. El uso de herramientas automáticas que ayuden al análisis de estos objetos podría incrementar la escala en la cual se pueden estudiar colecciones de datos de este tipo, facilitando nuevos e interesantes descubrimientos. En este proyecto, proponemos investigar métodos y algoritmos para poder realizar algunas de estas tareas de análisis sobre los gemelos digitales de los objetos originales.

En este proyecto, se definieron tres problemas principales para los cuales se investigaron algoritmos que pudiesen automatizar tareas relacionadas con el estudio de la cerámica antigua:

La digitalización [de cerámica antigua] permite la preservación de estos objetos para el futuro, y su estudio nos permite comprender las culturas del pasado.

- Dado un patrón en la superficie de una cerámica antigua, asignar etiquetas textuales que describan a dicho patrón.
- A partir de la imagen de la superficie de una cerámica antigua, identificar automáticamente las apariciones de todos los patrones relevantes pintados en la cerámica.
- A partir de una imagen o de un texto descriptivo de una cerámica antigua, encontrar el modelo 3D más relevante dentro de la colección digitalizada.

En las siguientes secciones de este artículo, describiremos cómo se abordó cada uno de estos problemas, cómo se propuso resolverlo utilizando técnicas de aprendizaje de máquina y de redes neuronales artificiales, cuáles son los principales resultados obtenidos en las evaluaciones de las técnicas implementadas, y cuáles son las principales conclusiones de este proyecto.

Etiquetado automático de patrones

El primer problema abordado fue el del etiquetado automático de patrones. Dada la imagen de un patrón, lo que se desea obtener son etiquetas textuales que describan el contenido del patrón. Por ejemplo, una imagen de un patrón que presenta varias líneas verticales organizadas en un cuadro horizontal podría tener como etiquetas textuales "líneas verticales" y "panel horizontal". Otro ejemplo es el que se muestra

Figura 1 / Patrón que aparece en una cerámica antigua de la colección del Museo Josefina Ramos de Cox en Lima, Perú [1]. Un ejemplo (ficticio) de etiqueta textual para este patrón es "escalera".

en la Figura 1, que muestra un patrón y una etiqueta textual (ficticia) asociada al patrón.

Para resolver esta tarea, nuestra hipótesis fue que esta se puede definir como un problema de clasificación multietiqueta: dado un conjunto de etiquetas textuales válidas, que corresponden a distintas "clases" posibles, se desea asignar todas las clases relevantes, que representan a las distintas etiquetas relevantes, a un patrón dado. Resolver esta tarea involucra tener una definición clara de cuáles son las etiquetas "válidas", y de distintos patrones de ejemplo en donde se sepa de antemano cuáles son las etiquetas relevantes para cada patrón de ejemplo. Para esto, tomamos como base la taxonomía propuesta por Norbert Kunisch en su libro "Ornamente Geometrischer Vasen: Ein Kompendium" [2]. En este libro, Kunisch describe más de 700 patrones distintos que pueden aparecer pintados en cerámica antigua, asignando un conjunto de etiquetas textuales por cada patrón descrito. Esta base es la que nos permitió implementar clasificadores multietiqueta, y así poder enseñarle al computador cómo describir textualmente un nuevo patrón.

Un primer enfoque para resolver el problema fue el de utilizar métodos clásicos de clasificación multietiqueta, como por ejemplo Binary Relevance [3], y métodos de la familia denominada "Extreme Multi-Class Labelling" [4]. Cabe destacar que ninguno de estos métodos fue capaz de resolver el problema con las etiquetas originales. La dificultad radica en que la distribución de estas etiquetas provenientes de la taxonomía de Kunisch es extremadamente desbalanceada: unas pocas etiquetas corresponden a una gran cantidad de patrones, y la gran mayoría de las etiquetas tiene muy pocos patrones asociados, lo que dificulta enormemente el poder entrenar un clasificador. Para resolver esto, se tuvo que realizar un preprocesamiento de las etiquetas para aliviar el problema del desbalanceo, de forma que una palabra represen-

Dada la imagen de un patrón, lo que se desea obtener son etiquetas textuales que describan el contenido del patrón.

tara a una etiqueta, y se eliminaron palabras comunes como artículos y conectores (palabras sin significado propio, que se conocen como *stop-words*). Con este conjunto modificado de etiquetas, fue posible entrenar los clasificadores multietiqueta y, en nuestra evaluación experimental, mostramos que es posible obtener al menos una etiqueta correcta para la mayoría de las imágenes de patrones [4, 5].

Un segundo enfoque fue utilizar técnicas de Inteligencia Artificial basadas en modelos multimodales, que son entrenados asociando texto e imagen. Es decir, estos modelos aprenden a asociar los textos con los patrones a los que corresponden. Al usar modelos multimodales como CLIP [6] o BLIP-2 [7], fue posible mejorar el resultado del etiquetado automático para el conjunto de etiquetas modificadas [8]. Por último, probamos con modelos multimodales más recientes, que ya no se basan en una tarea de clasificación, sino que generan directamente textos descriptivos a partir de una imagen. Con este último enfoque, se pudo obtener resultados iniciales prometedores usando el conjunto original de etiquetas [9], el cual hasta ahora había sido un problema intratable.

Detección de patrones en la superficie

El segundo problema abordado fue la detección automática de los patrones en la superficie de la cerámica antigua. Dada la imagen de la superficie, el problema consiste en detectar y marcar todas las apariciones de patrones relevantes. Para este problema, ocupamos una colección digitalizada de cerámica antigua perteneciente al Museo Josefina Ramos de Cox en Lima, Perú. Esta colección consiste en 82 modelos 3D de cerámica antigua, con motivos pintados en su superficie. Los patrones que aparecen en estas cerámicas tienden a ser repetitivos a lo largo del contorno de la cerámica, como en el ejemplo mostrado en la Figura 2. Para cada uno de estos objetos, un grupo de personas expertas anotaron manualmente todas las apariciones de patrones relevantes. Esta colección de referencia [1], de alta calidad, es la que nos permitió desarrollar y evaluar algoritmos automáticos de detección de los patrones, y poder evaluarlos con respecto a lo que haría una persona experta haciendo una examinación manual de la cerámica antigua.

Figura 2 / Ejemplo de aparición de un patrón en forma repetitiva, en la superficie de una cerámica antigua [1]. Note la anotación manual realizada por cada aparición del patrón, la que fue realizada por personas expertas en el dominio.

El segundo problema abordado fue la detección automática de los patrones en la superficie de la cerámica antigua.

Para resolver el problema, se evaluaron distintas técnicas de detección y segmentación de objetos. Primero se probaron técnicas del estado del arte como SegmentAnything [10], sin ningún tipo de modificación sobre el modelo original, pero estas técnicas no fueron capaces de realizar correctamente la detección de los patrones: sólo podían detectar unas pocas ocurrencias de estos motivos, y muchos quedaban sin ser detectados. Para superar este problema, se utilizaron técnicas de *fine-tuning* de modelos de redes neuronales artificiales, que permiten reentrenar estos modelos con ejemplos del tipo de patrones que se desea detectar, en este caso con imágenes provenientes de las cerámicas antiguas y las posiciones de todos los patrones relevantes que aparecen en ellas. Luego de este proceso de *fine-tuning*, y al repetir el experimento de detección de patrones, se lograron resultados satisfactorios con técnicas como YoloV8 [11] y RetinaNet [12]. Estos resultados muestran que, al menos con los modelos actuales, se requiere de una etapa de *fine-tuning* de los modelos con datos del contexto de la cerámica antigua, para que los modelos aprendan a detectar los patrones relevantes [13, 14].

Búsqueda en una colección digital de cerámica antigua

El tercer problema enfrentado fue el de la búsqueda dentro de la colección de modelos 3D de la cerámica antigua. Para este caso, la búsqueda que se requiere implementar es multimodal: 1) dado un texto describiendo un objeto, encontrar el modelo 3D de la cerámica antigua más relevante dentro de la colección, o 2) dada una imagen de un objeto, encontrar el modelo 3D de la cerámica antigua más relevante. Para poder resolver estas búsquedas multimodales, es necesario entrenar modelos de redes neuronales artificiales que aprendan en conjunto que cierta combinación de texto + imagen + modelo 3D corresponden a una misma cerámica. Es decir, al modelo se le enseña que estos tres datos distintos son correspondientes entre sí. Al entrenar los modelos de esta forma, luego se puede realizar la búsqueda ya sea a partir del texto o de la imagen de una cerámica, para que el sistema pueda encontrar el objeto 3D respectivo.

Como solución inicial para este problema, se propuso utilizar UNI3D [15], que es un modelo de red neuronal artificial que se puede entrenar utilizando las tres modalidades de datos simultáneamente (texto, imagen, y objeto 3D). Usando el conjunto de datos del Museo Josefina Ramos de Cox, se generaron textos descriptivos para las cerámicas antiguas de esta colección. Luego, se realizó un *fine-tuning* de UNI3D con estos datos. Finalmente, se utiliza UNI3D para, a partir de cada objeto de la colección, obtener representaciones de las distintas modalidades en el mismo "espacio de *embeddings*", es decir, se pueden representar las distintas modalidades de forma que sean comparables entre sí, lo que permite posteriormente implementar la búsqueda. Los resultados experimentales obtenidos muestran que esta metodología permite calcular las representaciones de los objetos y realizar las búsquedas en la colección de datos, pero es necesario que el sistema retorne varios objetos, ordenados en forma de ranking, para que haya una buena posibilidad que entre uno de ellos se encuentre el modelo 3D buscado. El principal problema detectado es que las representaciones comunes (el denominado "espacio de *embeddings*") no logran asociar completamente las tres modalidades, sino que se generan tres agrupaciones distintas por cada modalidad (un grupo para las descripciones de texto, un grupo para las descripciones de imágenes, y un grupo para las descripciones de los modelos 3D), lo que indica que el modelo no está aprendiendo correctamente cómo mezclar toda la información provenientes de los textos, imágenes y objetos 3D de la colección [16].

Conclusiones

En el Proyecto Fondecyt Regular 1230448 se investigaron métodos para etiquetar en forma automática patrones que aparecen en la superficie de cerámica antigua, para detectar en forma automática las posiciones en la superficie de la cerámica donde aparecen estos patrones, y para buscar dentro de una colección digitalizada de modelos 3D de cerámica antigua a partir de una descripción textual o de una imagen. Los problemas enfrentados son difíciles de resolver por variados motivos: 1) Los defectos presentes en la colección de datos, ya sea por partes faltantes, por erosión, o por errores durante el proceso de digitalización; 2) El desbalanceo de las etiquetas textuales (pocas etiquetas muy comunes y muchas etiquetas muy poco frecuentes); 3) La multimodalidad inherente de estos datos, que pueden ser representados como texto, imagen, u objeto 3D, que dificulta la implementación de buscadores que mezclen todas estas modalidades. A partir de los métodos desarrollados y de los resultados obtenidos de nuestras evaluaciones, llegamos a las siguientes conclusiones principales del proyecto: 1) Es necesario abordar el problema de desbalanceo de etiquetas para mejorar la eficacia de la clasificación multietiqueta, de forma de poder implementar un etiquetado automático; 2) Los modelos basados en redes neuronales artificiales provenientes del estado del arte no funcionan directamente sobre datos de este contexto, por lo que se requiere de un proceso previo de *fine-tuning* de estos modelos para que puedan lograr resultados eficaces en las tareas de etiquetado, detección de patrones y búsqueda dentro de la colección.

Si bien los resultados para cada uno de estos problemas muestran que es posible automatizar o semiautomatizar estas tareas, aún hay mucho espacio para mejorar la eficacia de las técnicas propuestas. En particular, la solución presentada para la búsqueda multimodal presenta problemas en cómo se están representando las distintas modalidades de los objetos, pero los resultados obtenidos permiten tener una línea de base sobre la cual empezar a mejorar. En este momento, estamos investigando técnicas de *relevance feedback*, que son técnicas que reciben retroalimentación de un usuario para indicar qué tan buenos (o malos) son los resultados que devuelve un sistema. En particular, estamos estudiando si es posible mejorar la detección de los patrones en la superficie de las cerámicas antiguas mediante la retroalimentación (lo que se conoce como "añadir al humano en el

ciclo"). Los resultados preliminares obtenidos muestran que es posible mejorar la eficacia del sistema a partir de la retroalimentación de la persona usuaria del sistema.

Por último, cabe destacar que varias Tesis de Magíster y Memorias de Ingeniería se realizaron en el contexto de este proyecto:

- **Matías Vergara:** tesis de Magíster en Data Science [4].
- **Sebastián Sepúlveda:** tesis de Magíster en Data Science [13].
- **Valentina Zúñiga:** memoria de Ingeniería Civil Eléctrica [8].
- **Juan Pablo Tacchi:** memoria de Ingeniería Civil en Computación [16].
- **Rodrigo Rapimán:** memoria de Ingeniería Civil en Computación [9].
- **Álvaro Garrido:** memoria de Ingeniería Civil en Computación (en curso).
- **Felipe Terraza:** tesis de Magíster en Tecnologías de la Información (en curso). ʙ

Agradecimientos

Este proyecto fue financiado por ANID FONDECYT Regular N° 1230448.

Referencias

[1] Stefan Lengauer, Iván Sipirán, Reinhold Preiner, Tobias Schreck, Benjamin Bustos. A benchmark dataset for repetitive pattern recognition on textured 3D surfaces. Computer Graphics Forum 40(5):1-8. Wiley & Sons Ltd., 2021.

[2] Norbert Kunisch. Ornamente Geometrischer Vasen: Ein Kompendium, 1998. https://www.zvab.com/buch-suchen/titel/ornamente-geometrischer-vasen/autor/kunisch/.

[3] Min–Ling Zhang, Yu-Kun Li, Xu-Ying Liu, Xin Geng. Binary relevance for multilabel learning: an overview. Frontiers Comput. Sci., 12(2):191{202, 2018.

[4] Matías Vergara. Aprendizaje Multietiqueta de Patrones Geométricos en Objetos de Herencia Cultural. Tesis de Magíster en Data Science. Universidad de Chile, 2023.

[5] Matías Vergara, Benjamin Bustos, Iván Sipirán, Tobias Schreck, Stefan Lengauer. Multi-Label Learning on Low Label Density Sets with Few Examples. Expert Systems With Applications 265:125942, 2025.

[6] Alec Radford et al. Learning transferable visual models from natural language supervision. En Actas International Conference on Machine Learning, pág. 8748–8763, 2021.

[7] Junnan Li, Dongxu Li, Silvio Savarese, Steven Hoi. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. En Actas International Conference on Machine Learning, pág. 19730–19742, 2023.

[8] Valentina Zúñiga. Asignación Automática de Etiquetas en Patrones de Herencia Cultural Utilizando Modelos de Lenguaje Multimodal. Memoria Ingeniería Civil Eléctrica. Universidad de Chile, 2025.

[9] Rodrigo Rapimán. Evaluación de Modelos Multimodales para el Etiquetado Automático de Patrones Geométricos en Objetos de Herencia Cultural. Memoria Ingeniería Civil en Computación. Universidad de Chile, 2025.

[10] Alexander Kirillov et al. Segment anything. arXiv preprint arXiv:2304.02643, 2023.

[11] Glenn Jocher, Ayush Chaurasia, Jing Qiu. YOLO by Ultralytics. 2023.

[12] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, Piotr Dollár. Focal loss for dense object detection. En Actas IEEE international Conference on Computer Vision, pág. 2980–2988, 2017.

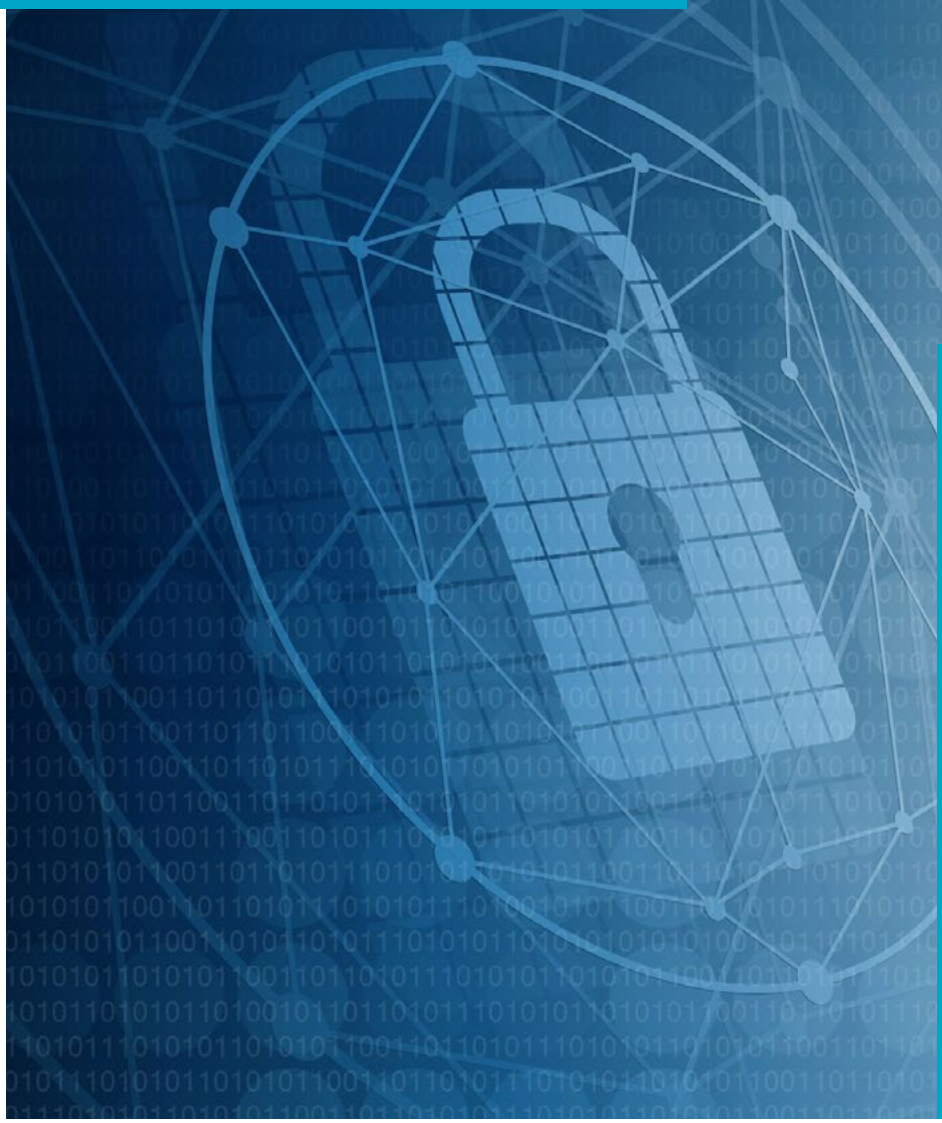
[13] Sebastián Sepúlveda. Reconocimiento de Patrones Repetitivos en Imágenes de Motivos de Herencia Cultural. Tesis de Magíster en Data Science. Universidad de Chile, 2024.

[14] Sebastián Sepúlveda, Benjamin Bustos, Iván Sipirán. Repetitive Patterns Recognition in Textures of Ancient Peruvian Pottery. ACM Journal on Computing and Cultural Heritage 17(4), Article 55, 2024.

[15] Junsheng Zhou, Jinsheng Wang, Baorui Ma, Yu-Shen Liu, Tiejun Huang, Xinlong Wang. Uni3D: Exploring Unified 3D Representation at Scale. En Actas 12th International Conference on Learning Representations (ICLR 2024), 2024.

[16] Juan Pablo Tacchi. Comparación de Modelos Multimodales Imagen+Texto+3D para la Búsqueda de Contenido Sobre Vasijas Arqueológicas. Memoria Ingeniería Civil en Computación. Universidad de Chile, 2025.

La Agencia Nacional de Ciberseguridad (o por qué la ciberseguridad no es un lujo)



Cristian Bravo Lillo

Ingeniero Civil en Computación por la Universidad de Chile y Ph.D. en Ingeniería y Políticas Públicas por la Carnegie Mellon University, especializado en seguridad usable. Actualmente es director del CSIRT Nacional (Equipo Nacional de Respuesta a Incidentes de Seguridad Informática) de la Agencia Nacional de Ciberseguridad.

crbravo@dcc.uchile.cl

Resumen / La recientemente creada Agencia Nacional de Ciberseguridad es un servicio público técnico con la misión de asesorar al presidente en la materia, colaborar con la protección de intereses nacionales, y coordinar instituciones de ciberseguridad. Su creación fue impulsada por incidentes graves, como la filtración del EMCO en 2022 y los ataques de ransomware a IFX Networks y GTD en 2023, que afectaron a servicios públicos esenciales.

En general, las personas y las organizaciones no se protegen contra riesgos de ciberseguridad pues son abstractos e inciertos; de manera similar, las personas frecuentemente no se protegen contra enfermedades potenciales en el futuro. Al igual que la salud, la ciberseguridad es una responsabilidad personal y del Estado. También de forma similar, el Estado no puede evitar que las personas se dañen a sí mismas (por ejemplo, usando malas claves), pero debe tomar medidas para prevenirlo. Prevenir incidentes es más efectivo que remediarlos.

La Agencia

La Agencia Nacional de Ciberseguridad (ANCI) es un servicio público de naturaleza altamente técnica. Su objeto es asesorar al presidente o presidenta en materias propias de ciberseguridad; colaborar en la protección de los intereses nacionales en el ciberespacio; coordinar el actuar de las instituciones con competencia en materia de ciberseguridad; velar por la protección, promoción y respeto del derecho a la seguridad informática; y coordinar y supervisar la acción de los organismos de la Administración del Estado en materia de ciberseguridad [1].

Una de las preocupaciones fundamentales de la Agencia es la ocurrencia de incidentes por ciberseguridad en el país. En 2022 y 2023 hubo en Chile casos graves de filtraciones de datos y de ransomware que afectaron al sector público, como la filtración masiva de correos electrónicos del Estado Mayor Conjunto (EMCO) en septiembre de 2022 [2], que obligó a la entonces Ministra de Defensa, Maya Fernández, a volver urgentemente al país para dar explicaciones y contener los efectos políticos de la filtración [3]; los ataques de ransomware sobre IFX Networks [4] y GTD en octubre de 2023 [5], que en el primer caso indisponibilizó el portal de compras públicas durante 9 días [6], y en el segundo caso afectó a más de 80 servicios públicos [7], entre ellos numerosas municipalidades, algunos servicios de salud y al menos dos servicios transversales muy importantes en el sector público: FirmaGob (https://firma.digital.gob.cl/) y DocDigital (https://doc.digital.gob.cl/). En octubre de 2023 el entonces proyecto de ley marco de ciberseguridad (que crea la ANCI) estaba siendo discutido en el Congreso, y los incidentes de ese mes galvanizaron el apoyo de los legisladores al proyecto y generaron un sentido de urgencia que de otra forma no hubiera tenido.

Sumado a lo anterior, el clima en Chile durante los meses que duró la elaboración del proyecto de ley estuvo teñido de una intensa preocupación por la seguridad física, que no hizo sino aumentar durante el gobierno del Presidente Boric. El proyecto de ley fue incluido en una "agenda corta" o agenda priorizada de medidas propuestas para mejorar la seguridad

En 2022 y 2023 hubo en Chile casos graves de filtraciones de datos y de ransomware que afectaron al sector público [incluyendo] el portal de compras públicas, numerosas municipalidades, algunos servicios de salud y dos servicios transversales muy importantes: FirmaGob y DocDigital.

en el país [8]. Esto sin duda benefició el avance del proyecto, que en otras condiciones no habría tenido el espacio legislativo necesario. Los proyectos que tienen una componente fuerte de tecnología rara vez tienen ese espacio legislativo. Todo esto probablemente contribuyó a que el proyecto de ley fuera aprobado y promulgado en tiempo récord. El primer proyecto de ley fue presentado al Senado por el gobierno del Presidente Piñera el 2 de marzo de 2022; una ley con contenido sustancialmente mejorado fue promulgada durante el gobierno del Presidente Boric el 26 de marzo de 2024: apenas dos años y 24 días después [9].

¿De qué se ocupa la Agencia?

En Latinoamérica, la mayor parte de los países tenemos problemas en apariencia mucho más básicos que la ciberseguridad. Según un estudio sobre pobreza multidimensional (que mide no sólo ingreso, sino también salud, vivienda, educación y empleo) presentado por CEPAL y PNUD en abril de este año, alrededor de una cuarta parte de la población de latinoamérica es pobre [10]. Por mucho que nos inquieten escenarios como los de Ejército, ChileCompras o GTD, preocuparse por este tipo de incidentes podría parecer un lujo que sólo pueden darse los países desarrollados, o en vías de desarrollo. Sin embargo, es falso creer que hoy uno puede

preocuparse de una cosa y no de la otra: la mayor parte de las necesidades básicas (agua y alcantarillado, electricidad, transporte, comunicaciones) dependen de Internet para ser provistas. De los factores que mide la pobreza multidimensional, al menos tres (salud, educación y empleo) dependen directa y fuertemente de Internet.

Si una cañería de agua se rompe por mala mantención y deja sin agua a una comunidad, normalmente uno no pensaría en este como un problema de ciberseguridad. Sí lo sería si alguien accediera sin permiso al sistema de control industrial que controla la cantidad de cloro que se agrega al agua para potabilizarla, y tratara de aumentar el nivel de cloro por sobre el límite saludable para la vida humana. Esto, por cierto, ocurrió en 2021 en Florida, Estados Unidos.[11]

Otro ejemplo: en 1999, en el condado de Maroochy, Australia, una persona tuvo una pelea con su jefe y renunció a su trabajo. Pidió al consejo del condado que lo recontrataran para otro rol, lo que no ocurrió. Enojado por esto, intervino el sistema SCADA que controlaba las 142 bombas de drenaje encargadas de llevarse las aguas servidas. Como resultado, las bombas dejaron de generar alarmas, y dejaron de bombear las aguas de desecho, inundando el lugar con más de 750 mil litros de aguas servidas. Los ríos se tiñeron de negro, muchos peces y fauna local murieron, y la población tuvo que soportar la pestilencia de las aguas servidas por semanas [12]. Hoy este es un caso de estudio para la comunidad técnica [13].

Tal como en los casos anteriores, muchos problemas hoy tienen aspectos de ciberseguridad de los que hay que hacerse cargo. Frente a un apagón como el del 25 de febrero de 2025 [14], tenemos que preguntarnos si la causa es la inundación de una planta de generación, un terremoto que botó parte de las torres de transmisión, o un ciberataque, como los que ocurrieron en Ucrania días antes de las navidades de 2015 [15] y 2016 [16]. El día del apagón, parte del equipo nos quedamos en las oficinas atentos a cualquier antecedente que indicara que se trataba de un ciberataque o incidente de ciberseguridad. Si así hubiera sido, habríamos juntado un equipo de analistas para identificar qué servidores o aplicaciones fallaron, dónde estaban físicamente, llamar a las personas encargadas, consultar por el estado de los servidores, y determinar desde ahí cuál era el mejor camino de acción (lo que podría haber incluido ir físicamente donde estuvieran los servidores que habilitaban el servicio, pedir u obtener los logs de las máquinas o de los dispositivos de red, analizarlos y llegar a nuestras propias conclusiones). Afortunadamente nada de eso fue necesario.

El proyecto de ley [que creaba la Agencia Nacional de Ciberseguridad] fue aprobado y promulgado en tiempo récord.

Nuestro ciberespacio es complejo

Nuestro ciberespacio es un sistema complejo; y los sistemas de este tipo tienden a fallar de maneras complejas y difíciles de prever. El aumento de la complejidad de este sistema, unido a una migración acelerada de servicios y procesos hacia el sistema, genera dos efectos: una dependencia creciente de nuestra sociedad de la red, y una incluso mayor complejidad que dificulta aún más el prever qué puede fallar.

Las instituciones que proveen Internet, que desde 2024 es considerado un servicio básico en Chile [17], se han transformado rápidamente en esenciales para la organización y provisión de todo el resto de los servicios, y para la coordinación de procesos, elaboración de productos, y un largo etcétera. Internet se ha transformado en una tecnología de base que posibilita o potencia todo lo demás.

Las instituciones que proveen servicios básicos, como el agua potable, la electricidad y el gas licuado o natural, son simplemente vitales y dependen mutuamente entre sí para su provisión: las instalaciones de agua potable dependen de la electricidad y de Internet para proveer agua continuamente; las plantas hidroeléctricas no sólo dependen del agua para generar electricidad: la producción y distribución de energía eléctrica depende de personas que requieren de agua de forma continua; etc. No es difícil imaginar escenarios catastróficos que comienzan con un corte de agua, electricidad o gas total o parcial en una ciudad o comuna.

Las instituciones públicas (que proveen un servicio esencial: la administración del Estado) son las que manejan más información de personas en el país. Por ejemplo, el SII tiene información sensible de personas que en la práctica alcanza a la totalidad de los mayores de 18 años. El Ministerio de Desarrollo Social administra y cautela información sensible de personas que alcanza al 90% del país. Si sólo una pequeña parte de esta información fuera filtrada, modificada o eliminada, generaría un perjuicio enorme a la sociedad.

¿De dónde puede venir el próximo gran problema en la red? Muchas veces los problemas son iniciados (o agravados) por personas con y (más frecuentemente) sin malas intenciones. Tal como Bomberos no puede evitar completamente que personas descuidadas inicien enormes incendios a partir de fogatas pequeñas, la Agencia no puede evitar que el administrador de un sistema SCADA que controla la potabilización de agua tenga "123456" como contraseña de la cuenta principal de administración. Sin embargo, sí puede brindar capacitación a las organizaciones sobre la importancia del uso de buenas contraseñas.

Lo anterior es clave: los seres humanos creamos, formamos parte, e influimos fuertemente en este entramado complejo que llamamos ciberespacio. Y la conducta humana es notoriamente difícil de predecir.

Un riesgo, por definición, implica protegerse contra algo que aún no ha ocurrido. Como los seres humanos preferimos las pérdidas inciertas (p. ej., invertir tiempo algún día en el futuro para aprender a usar un gestor de claves) sobre las pérdidas ciertas (p. ej., aprender a usar un gestor de claves hoy) [18], esto implica que nuestra tendencia natural es a nunca hacernos cargo hoy de un riesgo de ciberseguridad.

La mayor parte de las personas y las organizaciones no está dispuesta a invertir en protegerse, al menos no contra algo tan intangible y abstracto como un ciberataque. Por eso es necesaria una ley: porque el bienestar de todos depende de que las empresas y personas se protejan contra algo de lo que naturalmente no se protegerían.

En los casi 10 meses que llevamos en la ANCI, poco más de un 44% de los incidentes de efecto significativo parten con un compromiso de cuentas. Las medidas usuales que recomendamos para disminuir la probabilidad de pérdida son el uso de gestores de claves y de factores múltiples de autenticación (MFA), el bloqueo geolocalizado de conexiones remotas (VPN), y la configuración de dispositivos de seguridad perimetral para bloquear por defecto, en vez de permitir por defecto. Estimamos que estas medidas por sí mismas podrían detener casi un 70% de los ataques. En la mayor parte de las organizaciones, hemos observado cómo se sobreestima la importancia del antivirus y de los "aparatos" (firewalls, SIEMs, etc.), en detrimento de las medidas que realmente tienen importancia, como tener buenas claves (recordables pero no adivinables).

¿Qué recomendaciones se le puede dar a las organizaciones que quieren mejorar su nivel de seguridad, y cumplir con la ley marco?

La Agencia no puede evitar que el administrador de un sistema [...] tenga "123456" como contraseña [...]. Sin embargo, sí puede brindar capacitación [...] sobre la importancia del uso de buenas contraseñas.

1. **Reporta tus incidentes de ciberseguridad.** Sufrir un incidente no es algo malo: le sucede a todo el mundo, tarde o temprano; como un resfrío o un choque en auto. Hay personas a las que nunca les pasa, pero nadie miraría mal a alguien porque se resfría. Hay que dejar de pensar que es algo que nos muestra débiles, descuidados o ignorantes. Reportar incidentes tiene varias ventajas: si no tienes dinero para contratar a alguien, o no tienes a nadie que sepa algo sobre ciberseguridad, si reportas es probable que el CSIRT Nacional pueda ayudarte (depende de cuánto trabajo tengan en el momento).

   Existe la creencia de que luego de reportar llegarán las multas de la Agencia. A pesar de que cada caso debe analizarse en su propio mérito, hoy existe una tendencia fuerte a no multar a los que reportan. Hoy es muy difícil esconder que se ha sufrido un incidente: es por tanto un riesgo no reportar. Por el contrario: la tendencia es a fiscalizar a las instituciones que evidentemente han sufrido incidentes y no los han reportado. Ante la duda sobre si un incidente debe o no reportarse, es mejor reportar.

2. **¿Prestas un servicio esencial?** Los servicios esenciales están descritos en el artículo 4 de la ley: electricidad, transporte, agua potable, telecomunicaciones, etc. Si se provee un servicio esencial, las obligaciones que hay que cumplir son esencialmente dos: aplicar de manera permanente las medidas para prevenir, reportar y resolver incidentes de ciberseguridad; y reportar los incidentes de ciberseguridad. Ambas están pensadas para ser cumplidas por todo el mundo. Si crees o sabes que no puedes cumplir alguna, acude al CSIRT Nacional por ayuda.

3. **¿Fuiste nombrado un Operador de Importancia Vital (OIV)?** Los OIV están definidos en el artículo 5 de la ley. En esencia, se trata de aquellos prestadores de servicios esenciales que son considerados tan importantes, que su operación no puede ser detenida por un incidente de ciberseguridad. Ser OIV no es malo. La ley fue diseñada para permitirle a instituciones de distintos niveles de recursos cumplir con ella. Es fundamental tener una persona responsable de las medidas que hay que cumplir. No importa que esta persona no sea experta: puedes optar por el entrenamiento que brinda el CSIRT para cumplir con las obligaciones que la ley impone.

Algunas organizaciones han usado el hecho de haber sido nombradas OIV como un reconocimiento explícito a su importancia para la sociedad. Si vas a invertir tiempo y esfuerzo en esto, ¡haz que al menos valga la pena en términos de marketing!

La ciberseguridad y la salud

La ciberseguridad es ante todo una responsabilidad personal. Una analogía entre ciberseguridad y salud puede ser útil: para ser adultos funcionales, todas las personas tenemos que tener ciertas nociones de higiene básica, independientemente de que la mayor parte de nosotros no seamos profesionales de la salud. Tenemos que saber reconocer cuando estamos enfermos (cuando hay fiebre, por ejemplo), y cuando requerimos de la ayuda de un profesional de la salud. La misión del Estado en esto es asegurarle a todos la posibilidad de obtener prestaciones de salud. Para ello, debe asegurar que exista infraestructura pública de salud (hospitales y clínicas), y que los prestadores de salud, tanto públicos como privados, sigan las normas establecidas. A pesar de que el Estado no puede prohibir a alguien que se haga alcohólico tomando en exceso todos los días, o que coma grasas y azúcares en exceso, sabemos que una persona alcohólica, o con diabetes o dislipidemia no sólo se daña a sí misma y a su familia, sino al resto de la sociedad. Una persona enferma requerirá de tratamiento médico, medicamentos, terapias físicas, o tal vez de todas las anteriores. Utilizará servicios escasos y caros en una sociedad, y le quitará a otra persona la posibilidad de usarlos.

De manera similar, todas las personas tenemos que saber hacernos cargo de nuestro propio bienestar en términos de ciberseguridad. La misión de la Agencia es asegurarse de que la infraestructura más importante del país esté bien protegida en términos de ciberseguridad. La Agencia no puede prohibir a las personas que se hagan daño a sí mismas; por ejemplo, usando malas claves y haciendo sencillo el que otras personas sean capaces de adivinar esas claves, capturar sus cuentas y hacerles pasar un muy mal rato. Sin embargo, sí puede obligar a las instituciones que ofrecen servicios autenticados a cumplir con ciertas normas; por ejemplo, con un segundo factor de autenticación para que sea más difícil para un criminal tener acceso a las cuentas, incluso si las personas siguen usando malas claves.

En ciberseguridad, tal como en salud, hay algunos problemas que es posible prevenir (p. ej., que adivinen o exfiltren

credenciales de usuario), y algunos que sólo es posible remediar (p. ej., las denegaciones de servicio distribuidas). Prevenir un problema es usualmente más efectivo que lidiar con las consecuencias. Una forma práctica de prevenir incidentes es monitoreando el tráfico entrante (¡y saliente!) de los servicios públicos en búsqueda de patrones de tráfico maliciosos. Esto permite disminuir la cantidad de formas en que agentes de amenaza pueden afectar nuestros activos, y (en algunos casos) disminuir también el impacto sobre nuestros activos si estos llegan a ser afectados. B

Referencias

[1] Ley 21.663 (08/04/2024). Ley Marco de Ciberseguridad, Art. 10. https://bcn.cl/3isi2.

[2] Sepúlveda, N. (22/09/2022). Hackeo masivo al Estado Mayor Conjunto expuso miles de documentos de áreas sensibles de la defensa. Ciper Chile. https://www.ciperchile.cl/2022/09/22/hackeo-masivo-al-estado-mayor-conjunto-expuso-miles-de-documentos-de-areas-sensibles-de-la-defensa/.

[3] Ver por ejemplo: Caro, I., Gómez, R. (21/09/2022). Ministra Maya Fernández regresa de urgencia al país desde Nueva York tras hackeo a correos de las Fuerzas Armadas. La Tercera. https://www.latercera.com/politica/noticia/ministra-fernandez-interrumpe-gira-en-la-onu-y-regresa-de-urgencia-al-pais-tras-hackeo-a-correos-de-las-fuerzas-armadas/EQZJS4W2PVCHTL3FPOHTWN6FTY/; y Díaz, F. (21/09/2022). Ministra de Defensa retorna de emergencia a Chile tras hackeo a correos del Estado Mayor Conjunto. Biobio Chile. https://www.biobiochile.cl/noticias/nacional/chile/2022/09/21/ministra-de-defensa-deja-ny-y-vuelve-de-emergencia-a-chile-tras-hackeo-a-correos-de-fuerzas-armadas.shtml.

[4] Ver la alerta del CSIRT de Gobierno del 12/09/2023 (https://csirt.gob.cl/alertas/10cnd23-00108-01/), y el artículo de INCIBE del 06/10/2023 (https://www.incibe.es/incibe-cert/publicaciones/bitacora-de-seguridad/ciberataque-ransomware-contra-ifx-networks).

[5] El Mostrador (24/10/2023). GTD: reportan ciberataque a empresa de telecomunicaciones que ofrece servicios al Gobierno. El Mostrador. https://www.elmostrador.cl/noticias/pais/2023/10/24/ciberataque-a-empresa-de-telecomunicaciones-gtd-gobierno-reporta-servicios-publicos-afectados/.

[6] Fossa, L. (22/09/2023). Lo que se sabe y lo que no del ciberataque que afectó a Mercado Público en Chile. Interferencia. https://interferencia.cl/articulos/lo-que-se-sabe-y-lo-que-no-del-ciberataque-que-afecto-mercado-publico-en-chile.

[7] Cárdenas, L. (14/11/2023). Agencia de ciberseguridad del gobierno califica ataque a GTD como un "incidente grave y masivo" y Subtel alista reunión con gerentes. La Tercera. https://www.latercera.com/pulso-pm/noticia/agencia-de-ciberseguridad-del-gobierno-califica-ataque-a-gtd-como-un-incidente-grave-y-masivo-y-subtel-alista-reunion-con-gerentes/GUQERJ7OJBH6NOSUDG76M7DEQE/.

[8] Riquelme, I. (21/02/2025). Leyes y proyectos de ley pertenecientes a la agenda de seguridad del Gobierno del Presidente Gabriel Boric Font. Biblioteca del Congreso Nacional de Chile. https://obtienearchivo.bcn.cl/obtienearchivo?id=repositorio/10221/36983/1/BCNSeleccionPL_AgendaSeguridad.pdf.

[9] Biblioteca del Congreso Nacional (08/04/2024). Historia de la Ley N° 21.663. https://www.bcn.cl/historiadelaley/nc/historia-de-la-ley/8286/.

[10] CEPAL (02/04/2025). CEPAL y PNUD presentan un nuevo Índice de Pobreza Multidimensional para América Latina. Sitio web CEPAL. https://www.cepal.org/es/noticias/cepal-pnud-presentan-un-nuevo-indice-pobreza-multidimensional-america-latina.

[11] BBC (08/02/2021). Hacker tries to poison water supply of Florida city. Sitio web de BBC. https://www.bbc.co.uk/news/world-us-canada-55989843.

[12] Levi, R. (07/02/2016). What the Maroochy Incident taught us about Cyber Warfare. Medium. https://medium.com/curious-minds/what-the-maroochy-incident-taught-us-about-cyber-warfare-4a1abd6abcfc.

[13] Ver, por ejemplo: Abrams, M., Weiss, J. Malicious Control System Cyber Security Attack Case Study–Maroochy Water Services, Australia. Case #08-1145, MITRE. https://www.mitre.org/sites/default/files/pdf/08_1145.pdf; y Sayfayn, N., Madnick, S. (mayo de 2017). Cybersafety Analysis of the Maroochy Shire Sewage Spill. Working Paper CISL #2017-09. MIT Sloan School of Management. https://web.mit.edu/smadnick/www/wp/2017-09.pdf.

[14] Wikipedia (español), artículo "Apagón de Chile de 2025". https://es.wikipedia.org/wiki/Apag%C3%B3n_de_Chile_de_2025.

[15] Wikipedia (inglés), artículo "2015 Ukraine power grid hack". https://en.wikipedia.org/wiki/2015_Ukraine_power_grid_hack.

[16] Wikipedia (inglés), artículo "2016 Kyiv cyberattack". https://en.wikipedia.org/wiki/2016_Kyiv_cyberattack.

[17] Ley 21.678 (03/07/2024). Establece el acceso a Internet como servicio público de telecomunicaciones. https://bcn.cl/3kr7q.

[18] Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. Econometrica, 47(2), 363-391.

Haciendo doble-click sobre la Ley Marco de Ciberseguridad:

Motivaciones, desafíos y oportunidades

Eduardo Godoy Vega

Magíster en Docencia para la Educación Superior por la Universidad Andrés Bello e Ingeniero Civil en Computación por la Universidad De Chile. Profesor Adjunto del Departamento de Ciencias de la Computación de la Universidad de Chile. Además, es consultor senior en seguridad de la información y ciberseguridad, implementador líder y auditor de la norma ISO27001 al frente de su empresa, CISOVirtual, además de perito judicial en temas relacionados.

egodoy@dcc.uchile.cl

Resumen / La existencia de una ley de ciberseguridad es fundamental porque los ataques informáticos representan hoy un riesgo real para el funcionamiento del Estado, las empresas y la vida cotidiana. Incidentes recientes —como el ocurrido en la Subsecretaría de Prevención del Delito en septiembre de 2025— evidencian que el país requiere una estructura formal para coordinar su protección digital.

Aunque podría pensarse que basta con el "sentido común", este no es suficiente: la industria tecnológica ha generado una falsa sensación de seguridad al afirmar que ciertos servicios, como la nube o ciertos teléfonos, son "muy seguros". Esto ha llevado a que muchas personas e instituciones subestimen los riesgos reales.

La Ley Marco de Ciberseguridad (Ley 21.663), promulgada en 2024, crea la Agencia Nacional de Ciberseguridad (ANCI) y define obligaciones, estándares y protocolos para los sectores críticos —como electricidad, agua, telecomunicaciones, finanzas y transporte— con el fin de que operen de forma segura y continua. Esta normativa invita a todos los actores del Estado, incluidas empresas y ciudadanos, a elevar su conocimiento y conciencia sobre los riesgos digitales, entendiendo que los atacantes actuales son organizaciones criminales complejas y no simples aficionados.

Para las ingenieras y los ingenieros en computación, la ley abre oportunidades laborales en áreas como cumplimiento regulatorio, análisis de riesgos, aseguramiento de la continuidad operativa, respuesta a incidentes, consultoría y desarrollo de tecnologías seguras.

Introducción

¿Por qué un país requiere una ley de ciberseguridad?, ¿no debería bastar con el "sentido común"?, ¿qué oportunidades presenta esta ley para las ingenieras y los ingenieros en computación? En este artículo pretendo dar algunas luces y sembrar algunas dudas en el lector.

Demos algo de contexto... Según el reporte de riesgos globales del Foro Económico Mundial 2025 [1], temas relacionados con ciberespionaje y guerra cibernética están dentro de los *top 10* riesgos con mayor impacto que podrían afectar gobiernos y estabilidad social, entre otros.

Entonces, el estado debe tomar medidas para protegerse frente a ataques cibernéticos que puedan poner en riesgo su continuidad operativa. Acá debemos entender, según Carré de Malberg [2], que define al Estado como "una comunidad humana, fijada sobre un territorio propio, que posee una organización que resulta para ese grupo, en lo que respecta a las relaciones con sus miembros, una potencia suprema de acción, de mando y coerción", entonces, como Estado, es decir todos los actores que participamos de él, debemos tomar acciones concretas en la ciberseguridad; de ahí la importancia de la Ley Marco de Ciberseguridad, ya que define criterios y prioridades al momento de definir qué, quiénes y cómo deben coordinarse estas acciones.

Hago hincapié en el mensaje de "todos los actores". Aún es común escuchar en empresas e instituciones que la ciberseguridad es un problema del área TI, cuando en realidad es un problema de *continuidad de negocio*, por lo tanto es un problema o preocupación transversal a toda la institución o empresa. Una falla en ciberseguridad te puede dejar fuera de operación o incluso puede significar el cierre de las operaciones.

El 11 de septiembre de 2025 leíamos esto en los medios de comunicación:

> *La Subsecretaría de Prevención del Delito (SPD), institución liderada por Carolina Leitao, registró un "incidente informático" que afectó a equipos institucionales y está generando intermitencia en servicios virtuales.*[1]

Espero haber dejado meridianamente claro el porqué necesitamos de esta ley.

¿Y por qué el sentido común no es suficiente?

Interesante pregunta, si tengo una actividad que depende fuertemente de la computación, debería ser obvio que debo preocuparme de cuidarla, pero acá aparecen algunos vicios. Para bien o para mal muchas veces nosotros mismos, los ingenieros, damos certezas de cosas que no lo son —entre ellas, "no se preocupe, la nube es *muy* segura", delegando la seguridad en el proveedor de servicios de cómputo en la nube; otro ejemplo, "esa marca de celular es inviolable, es *muy* segura", sin revisar que en la lista de Apps con algún tipo de *malware* también aparecen listados. Es así como hemos sembrado, entre los técnicos y el marketing, una falsa sensación de seguridad y eso es muy malo; es peor sentirse seguro, cuando realmente no lo está, a tener conciencia de las inseguridades o vulnerabilidades que poseo.

Entonces, a ese "sentido común" lo hemos ido bloqueando con siglas y tecnologías difíciles de explicar y aún más difíciles de entender para la gente que está fuera del rubro.

Resumen, no, no basta con el sentido común.

1 Fuente: Emol.com https://www.emol.com/noticias/Nacional/2025/09/11/1177644/equipos-institucionales-spd-incidente-informatico.html

Aún es común escuchar en empresas e instituciones que la ciberseguridad es un problema del área TI, cuando en realidad es un problema transversal a toda la institución o empresa.

Vamos a lo más concreto.... La Ley 21.663 es la Ley Marco de Ciberseguridad [3]; luego de una larga discusión iniciada en marzo de 2022, fue promulgada en abril de 2024 y coloca a Chile al día en cuanto regulación en este tema.

Crea la Agencia Nacional de Ciberseguridad (ANCI), que tiene por objetivo, entre otros, definir protocolos, estándares técnicos y regulaciones obligatorias para los servicios esenciales y Operadores de Importancia Vital; es decir, le dirá a los sectores que son fundamentales para que el país funcione cómo deben protegerse, existiendo el principio de proporcionalidad, es decir, no puede ser más cara la protección que lo protegido; ahora este tema toma tonos éticos en este punto, cuando lo que protegemos es la vida de un ser humano o los derechos de niños, niñas y adolescentes, eso tiene un valor infinito.

¿A qué nos obliga o invita la ley de ciberseguridad? Lo primero y más importante es a crear conciencia en el tema; todos los actores del estado, es decir, empresas, instituciones y por cierto los ciudadanos debemos incrementar nuestro saber en ciberseguridad. Acá un primer gran problema: para el común de las personas, si escuchan hablar de ciberseguridad, imaginan a un hacker —típicamente un o una joven poco sociable, que durante la noche y usando conocimientos de las artes oscuras, se mete a los computadores de las empresas para robar datos, modificar información e incluso apagar complejos sistemas informáticos— cuando sabemos que existen verdaderas empresas dedicadas a este delito... son organizaciones criminales. Si tomamos conciencia que los atacantes no son esta parodia de joven en su dormitorio, rodeado de cajas de pizza que está intentando romper un sistema, sino que son verdaderas corporaciones, entonces tomaremos conciencia que para protegerse no basta con la buena voluntad o apostar a la suerte de "a mí no me pasará".

¿Cuáles son los servicios esenciales?

Corresponden a los siguientes:

- Generación, transmisión o distribución eléctrica.
- Transporte, almacenamiento o distribución de combustibles.
- Suministro de agua potable o saneamiento.
- Telecomunicaciones.
- Infraestructura digital, servicios digitales y servicios de tecnología de la información gestionados por terceros.
- Transporte terrestre, aéreo, ferroviario o marítimo, así como la operación de la infraestructura respectiva.
- Banca, servicios financieros y medios de pago.
- Administración de prestaciones de seguridad social.
- Servicios postales y de mensajería.
- Prestación institucional de salud (hospitales, clínicas, consultorios, centros médicos).
- Producción y/o investigación de productos farmacéuticos.

Si la empresa o institución está dentro de alguna de estas categorías, tienen la obligación de registrarse en la ANCI y de informar cada vez que sufran algún ataque de ciberseguridad que sea considerado relevante.

¿Qué pasa si la empresa no informa?

Quizás lo primero que debemos pensar es ¿por qué una empresa no querría informar? Generalmente es por el temor a que su imagen se vea dañada o incluso a potenciales efectos legales de ciudadanos u otras empresas que pudieran verse afectados por este ataque. Pues bien, en el caso de no hacerlo, la ANCI está facultada para cursar multas, que en el peor de los casos supera los 1.600 millones de pesos. Importante, el mal informar, es decir, disfrazar el evento por otro de manera consciente, también es constitutivo de falta; por ejemplo, sufrí el robo de mis datos y reporto que me hicieron un ataque de Denegación de Servicios Distribuido (DDoS).

Sin duda que para cada uno de estos sectores se nos pueden ocurrir ejemplos claros de empresas o instituciones que las entendemos como esenciales. Por ejemplo, en salud, podríamos pensar desde hospitales de alta complejidad hasta la atención primaria; en telecomunicaciones, todos los operadores de conectividad de datos y voz —de sólo pensar (o recordar) caídas, suspensiones o incluso intermitencias en sus servicios ya podemos ver el caos que eso generaría: no poder usar el banco, pagar el metro y un largo etcétera... la vida de muchos ciudadanos se vería alterada... no en vano somos un país con una alta adopción tecnológica. Ahora bien, dentro de estos sectores hay operadores que son más relevantes que otros, no es lo mismo que un pequeño prestador de transporte terrestre tenga una suspensión en sus servicios a que lo haga la principal línea área nacional o aquella que tiene ruta única, como puede ser Santiago–Rapa Nui. A ellos,

la ANCI los podrá nominar como Operadores de Importancia Vital (OIV); y con ese nombramiento vienen más responsabilidades y mayor control de sus cumplimientos.

Pero, ¿qué significa en la práctica ser un OIV?

Pues bien, al ser OIV, la ANCI tiene facultades para poder exigir algunos estándares más altos de protección. También contarán con la supervisión y ayuda de la ANCI frente a situaciones de ataques que puedan o estén colocando en riesgo la continuidad operacional. Y aunque no todo es garrote, hay que estar claros que, en el caso de los OIV, las multas se pueden llegar a duplicar con respecto a un servicio esencial.

Las empresas e instituciones que han aparecido en las primeras nóminas de OIV se lo han tomado de dos formas muy marcadas, algunas lo consideran un elogio, un reconocimiento al servicio que prestan y el valor que este tiene para el Estado; otras, lo han percibido como una mochila, que significará mayor inversión, aumento de costos de operación e incluso, que los podría dejar fuera de negocios al perder competitividad.

Muchas veces nosotros mismos, los ingenieros, damos certezas de cosas que no lo son: "no se preocupe, la nube es muy segura".

Aún falta ver cómo reacciona el ecosistema: si valora que una empresa potencialmente proveedora sea un OIV, o simplemente será un dato sin relevancia al momento de evaluar ofertas.

Ahora bien, ¿qué oportunidades trae esta ley para las ingenieras y los ingenieros en computación?

Las empresas deberán comenzar a implementar protocolos y estándares de ciberseguridad. Algunos sectores van más adelantados que otros. En el mundo de la energía eléctrica llevan 5 años (desde 2020) implementando medidas de ciberseguridad y definieron el estándar NERC-CIP para su infraestructura crítica; la banca, a través de regulaciones de la CMF, ha hecho lo propio; el mismo Estado en el pasado cercano

Las empresas e instituciones que han aparecido en las primeras nóminas de Operadores de Importancia Vital se lo han tomado de dos formas muy marcadas: algunas lo consideran un elogio; otras, lo han percibido como una mochila.

definió a ISO27001 como la norma de referencia y varios servicios, ministerios e instituciones avanzaron en la definición de políticas y procedimientos según dicta ese estándar. Lo que se espera es que a cada vertical se le encuentre el estándar que le sea más apropiado. Pero lo realmente interesante, es que estos estándares traen, generalmente, requerimientos para toda la cadena de proveedores de estos OIV, quienes deberán cumplir con las exigencias que sus clientes les pidan. Esto será interesante de observar, veremos negociaciones de contratos, revisiones de costos, etc. —pero sin duda, tendremos una mejora en la ciberseguridad de todo el ecosistema, sin importar si eres OIV, servicio esencial o simplemente un proveedor de alguno de ellos. Así que se van a necesitar ingenieros e ingenieras que sepan implementar estos estándares, implementar soluciones de ciberseguridad e ingenieros/as que sepan gobernar esta nueva realidad. A este último se le llama CISO (Chief Information Security Officer), que tiene por desafío principal hacer conversar las necesidades de la empresa, su estrategia con las soluciones de ciberseguridad que el mercado está ofreciendo.

Sin duda que esto recién comienza... La Ley Marco de Ciberseguridad no ha dejado a ningún actor de la industria indiferente y eso ya es un gran logro: hacer que se pregunten si realmente deben invertir en ciberseguridad, incluso preguntarse si son candidatos a ser OIV. El solo hecho de abrir la discusión, sacándola del área de TI y colocándola en la mesa de directorios, juntas de gerentes o en la preocupación del pequeño empresario es sin duda el primer gran aporte de esta ley. B

Referencias

[1] World Economic Forum (2025). *Global Risks Report 2025, 20th Edition*. https://reports.weforum.org/docs/WEF_Global_Risks_Report_2025.pdf.

[2] Carré de Malberg, R. (1998). *Teoría general del estado* (J. Ferrándiz & A. Orihuela, Trads.). Fondo de Cultura Económica.

[3] Biblioteca del Congreso Nacional de Chile. Ley 21663: Ley Marco de Ciberseguridad. https://www.bcn.cl/leychile/navegar?idNorma=1202434.

Consideraciones para la aplicación de la nueva Ley de Protección de Datos Personales:

Un puente entre la regulación y la arquitectura de datos

Fernanda Carvajal

Abogada por la Universidad de Chile y Máster en Ciberseguridad por la Universidad de Barcelona. Se desempeña como Asociada en el área de protección de datos, ciberseguridad e IA del estudio jurídico Prieto Abogados, asesorando a empresas nacionales y extranjeras en materias de cumplimiento regulatorio orientado principalmente a la gestión de información y al desarrollo de iniciativas tecnológicas, investigando, levantando procesos, desarrollando estrategias y elaborando la documentación necesaria.

Fernanda Carvajal Gezan

Resumen / He escrito este artículo desde mi rol como abogada, una con profundo interés en la tecnología y su impacto en las personas. Lo he escrito a partir de mi trabajo cotidiano asesorando e implementando el nuevo marco regulatorio para el tratamiento de datos personales y la seguridad de la información en distintas industrias, y lo he preparado como alguien que disfruta lo que hace, pero que debe enfrentar las dificultades inherentes a la coordinación y comunicación entre las diversas áreas de las organizaciones que, más allá de tratar datos, tienen, cada una, su propio lenguaje, enfoque y prioridades. Este texto no busca una exposición jurídica de la nueva ley, sino una presentación con enfoque práctico de los principales aspectos que afectan a las áreas de gestión de la información y tecnología, que son las llamadas a convertir en realidad el cumplimiento normativo y a materializar lo que los abogados esbozamos muchas veces como directrices, criterios y principios.

La computación y la programación son disciplinas relativamente nuevas que se han caracterizado, en buena medida, por el asombro y distancia que producen debido a su dificultad conceptual y lógica, que les han significado un importante podio en instituciones y empresas, pero cierto aislamiento en cuanto a la comprensión de los fenómenos involucrados, la toma de decisiones y la presentación de conclusiones, donde muchas veces basta con constatar ante gerencias y directorios que no han ocurrido incidentes relevantes, cuando lo que hay de fondo es la gestión de una compleja infraestructura que parece invisible para el resto de la organización.

Este escenario ha tenido ventajas y desventajas: la frustración frente a la falta de comprensión sobre los límites y fundamentos del trabajo técnico, pero también la relativa independencia que este desconocimiento permitía en el quehacer cotidiano.

Y, aunque las áreas de seguridad de la información y ciberseguridad han impulsado sostenidamente la adopción de mejores prácticas, la concientización sobre riesgos y el robustecimiento de procesos, el contexto en que estas iniciativas eran sólo automotivadas y privativas de estas áreas parece llegar a su fin con la aprobación en Chile de las Leyes 21.719 sobre Protección de Datos Personales y 21.663, conocida como Ley Marco de Ciberseguridad.

En vista de ello, este artículo persigue un acercamiento práctico a la nueva Ley de Protección de Datos Personales, partiendo con un breve contexto de la norma y la presentación de algunos conceptos relevantes que deberán considerarse en los mapas de procesos de las organizaciones, para seguir con la forma en que dichos conceptos deben ser aplicados en casos de uso. También se abordan los principales elementos del gobierno de datos que deberán considerarse en los planes de implementación del nuevo marco regulatorio.

El cambio regulatorio

A pesar de que Chile cuenta con una ley que regula el tratamiento de los datos personales desde el año 1999, habiendo sido pionero en la región al normar esta materia mediante la Ley 19.628, esta norma, lamentablemente, tuvo falencias estructurales que la hicieron prácticamente inaplicable. Entre ellas destacan la ausencia de sanciones significativas, la falta de una autoridad reguladora que interpretara y fiscalizara su cumplimiento, y un procedimiento excesivamente largo y costoso para ejercer los derechos. De hecho, ante la negativa o el silencio de una institución, el titular debe contratar a un abogado y recurrir a los Tribunales Ordinarios de Justicia incluso para saber qué datos suyos mantiene una empresa o para solicitar su eliminación.

Este escenario cambia de forma significativa con la Ley 21.719, publicada en diciembre de 2024, que modificará la Ley 19.628 a partir del 1 de diciembre de 2026, fecha en que entrará en vigencia. La nueva normativa crea la Agencia de Protección de Datos Personales, facultada para interpretar las reglas de tratamiento, fiscalizar su cumplimiento y aplicar sanciones. Para ello se establece un catálogo de 31 infracciones, clasificadas como leves, graves y gravísimas, que pueden dar lugar a multas de hasta 20.000 UTM, con incrementos en caso de reincidencia.

Asimismo, la ley incorpora un conjunto de derechos para los titulares y un procedimiento más simple y expedito para ejercitarlos directamente. También establece un esquema claro de obligaciones y responsabilidades para todas las personas y entidades que traten datos personales fuera del ámbito doméstico.

Conceptos clave para el cumplimiento

Lo primero es comprender que los datos personales se refieren únicamente a aquellos que identifican personas naturales y que su tratamiento contempla todo tipo de operaciones sobre ellos, desde su captura, análisis, transferencia, procesamiento, eliminación o su mero almacenamiento. Esto último es relevante para quienes prestan o utilizan servicios de nube para el respaldo, redundancia o procesamiento de datos, ya que si bien es cierto que los prestadores no conocen el contenido de la información, sí asumen obligaciones respecto de las actividades de tratamiento de datos que realizan, desde su recepción, almacenamiento, seguridad, envío y eliminación, según corresponda.

Otro elemento a destacar es que, pese a la noción habitual de que los datos relativos a personas se entienden como "sensibles" dentro de la organización, esta ley establece una clasificación diferente, señalando que son datos sensibles aquellos

El tratamiento [de datos personales] contempla todo tipo de operaciones sobre ellos, desde su captura, análisis, transferencia, procesamiento, eliminación o su mero almacenamiento.

que refieren características físicas o morales de las personas, hechos o circunstancias de su vida privada o intimidad, que revelan el origen étnico o racial, la afiliación política, sindical o gremial, la situación socioeconómica, las convicciones ideológicas, filosóficas o creencias religiosas, los datos relativos a la salud, al perfil biológico humano, los datos biométricos, y la información relativa a la vida u orientación sexual, o a la identidad de género de una persona. Adicionalmente, se incluyen como "categoría especial" los datos de menores de edad, los datos de geolocalización y otros referidos a la realización de estudios o investigaciones. Tanto los datos sensibles como estas categorías especiales, que pueden referirse conjuntamente como "datos especialmente protegidos", tienen requisitos y/o resguardos adicionales para su tratamiento, los que deben tenerse en consideración para la creación de catálogos de datos, su trazabilidad, configuración de permisos y restricciones. Este factor también cobra importancia frente a incidentes o vulneraciones de seguridad que puedan afectar la información de personas, ya que cuando se comprometan datos personales será necesario reportar los incidentes a la Agencia de Protección de Datos Personales, a la Agencia Nacional de Ciberseguridad y, en algunos casos, también a los titulares afectados, siendo fundamental la capacidad de detectar a la brevedad si existe tal compromiso sobre la información de personas y de qué tipo de información se trata. La necesidad de identificar rápidamente si se ha afectado información de personas y si es necesario reportar a autoridades y/o titulares deberá incorporarse en los planes de respuesta a incidentes de las organizaciones, en los que podrán participar áreas legales y de comunicaciones cuando la gestión del incidente, sus efectos y obligaciones trascienda la esfera de control interno de la organización.

Responsables y encargados

En cuanto a las organizaciones que tratan datos personales, podemos encontrar entidades responsables y encargadas. Son responsables aquellas que deciden sobre los medios y fines de los tratamientos, mientras que actúan como encargados aquellos a los que un responsable ha mandatado determinados tratamientos que, en todo caso, se deben realizar por cuenta del responsable, que mantiene la responsabilidad frente a los titulares. Este esquema de responsabilidades vuelve determinantes las evaluaciones de seguridad de la información, estándares de protección de datos y auditorías periódicas a proveedores, especialmente a los que desempeñan funciones críticas o que tratan datos especialmente protegidos por cuenta del responsable, ya que una vulneración de seguridad del proveedor encargado, que exponga los datos personales que le han sido confiados por el responsable podrá derivar en sanciones para este último en caso de no haber adoptado medidas y controles robustos para garantizar que aquel proveedor cumpliera con estándares de tratamiento de datos adecuados. Así, la realización de evaluaciones previas a la contratación, la adopción de planes de mitigación super-

visados y, posteriormente, la realización de auditorías al cumplimiento de los estándares requeridos serán medidas fundamentales para el cumplimiento de esta norma.

Cabe destacar que el establecimiento de cláusulas contractuales como las que se acostumbran en la actualidad seguirá siendo necesario, pero ciertamente insuficiente para demostrar una debida diligencia. El objetivo es desarrollar medidas de seguridad y control proactivas, que den cuenta de la realidad de los proveedores, sus riesgos y fortalezas, evitando la aplicación de controles meramente declarativos que sólo podrían derivar en acciones reactivas e indemnizatorias entre las partes, ya que el foco se extiende también a la protección de los terceros cuyos datos son utilizados para el provecho de las empresas y otras instituciones.

Trazabilidad y gestión de derechos

Los derechos de los titulares son otro aspecto a gestionar, ya que las personas podrán consentir determinados tratamientos, debiendo quedar constancia documental o electrónica que permita su comprobación frente a reclamos o consultas por parte de la Agencia, pero, tal como este consentimiento puede otorgarse, puede ser revocado libremente y en cualquier momento, circunstancia que deberá gestionarse internamente para impedir la continuación de las actividades de tratamiento cuya autorización ha sido revocada.

Además de esta libertad de consentir, los titulares tendrán otros derechos, conocidos con la sigla ARSOP, que incluyen el *Acceso*, es decir, la posibilidad de conocer los datos que se tratan, su origen, fines y destinatarios, entre otros aspectos; *Rectificación* para corregir o actualizar los datos; *Supresión* para eliminar los datos que el titular no desea que sigan siendo tratados, siempre que se cumplan las causales legales; *Oposición* a determinados tratamientos, incluyendo la toma de decisiones automatizadas que produzcan efectos significativos y en las que no exista participación humana; *Portabilidad* para obtener del responsable una copia de los datos susceptible de ser compartida con otros responsables de datos. Dado que todos estos derechos pueden ser ejercidos directamente por los titulares y que la organización se encuentra obligada a responder dentro de 30 días, será necesario estructurar sistemas centrados en los titulares, que permitan localizar la información y asociarla, lo que requiere de su catalogación y normalización previa, del conocimiento de las fuentes en que se encuentra almacenada, su dueño funcional y sistemas. Además, todo el proceso de ejercicio de derechos debe quedar documentado y ser trazable respecto de las acciones y respuestas llevadas a cabo tanto por el titular como por la organización a modo de respuesta, lo que requiere de la identificación y autenticación del solicitante, la eventual validación de sus poderes, del registro de las solicitudes, coordinación de sistemas, control de plazos y mantenimiento de la evidencia de cumplimiento.

Conservación y eliminación

En cuanto al gobierno de los datos y su seguridad a lo largo de todo el ciclo de vida, será necesario que las áreas dueñas de los datos tengan claridad acerca de los datos que capturan, cuáles son sus fines y durante cuánto tiempo necesitarán mantener esta información, debiendo establecer mecanismos de supervisión respecto de plazos o criterios de conservación y eliminación de los datos. Esta ley exige transparentar en la política de privacidad los plazos para los que serán utilizados los distintos tipos de datos personales y las personas tendrán el derecho a solicitar su supresión en cualquier momento, sin perjuicio de que las organizaciones puedan hacer valer otros fundamentos, denominados "bases de licitud", para mantener los datos. De modo que se debe conocer qué se tiene, dónde se tiene, con qué justificación y para qué.

Hoy en día es bastante común que no exista un control riguroso sobre la información que se conserva, la cual suele mantenerse de manera indefinida con fines de respaldo o "por si acaso" pudiera ser necesaria en el futuro. Sin embargo, el establecimiento de políticas claras de conservación y eliminación de datos facilita significativamente el control sobre lo que se posee y permite justificar adecuadamente las decisiones respecto de la información que se decide eliminar. La higiene de la información también es una forma de seguridad.

Ahora bien, la eliminación de los datos deberá llevarse a cabo mediante los mecanismos seguros definidos por las instancias técnicas de cada organización. Estos pueden incluir métodos como la sobreescritura, la desmagnetización, la destrucción física o la anonimización. En este último caso, un dato que ya no permite la identificación de una persona deja de considerarse dato personal; no obstante, es fundamental que dicha disociación se realice de modo tal que la reidentificación no sea posible utilizando métodos o recursos razonables.

Además, será necesario garantizar la seguridad dentro de la operación, especialmente cuando se utilicen o envíen, interna o externamente, datos especialmente protegidos. En estos casos será menester aplicar medidas como la segregación de accesos basada en la necesidad de uso y en el perfil del cargo; el cifrado de la información, tanto en reposo como en tránsito; el enmascaramiento o la seudoanonimización cuando no resulte necesario conocer el dato en su forma original para cumplir con el propósito del tratamiento; la priorización del uso de datos sintéticos para ambientes de pruebas; entre otras medidas equivalentes.

Flujo transfronterizo de datos

Otro aspecto sobre el que se deberá tener claridad corresponde a los flujos transfronterizos de datos personales, ya que constituye una obligación de los responsables transparentar si es que la organización envía los datos al extranjero, ya sea en forma directa o a través de terceros, y adoptar garantías adecuadas para la seguridad de los datos.

Para cumplir con esta obligación será necesario identificar si la operación y/o los servicios que se ofrecen necesitan de este tipo de flujo, lo que puede ocurrir cuando se contacta o requiere de un operador foráneo, cuando se ofrecen productos suministrados por proveedores extranjeros o, sencillamente, cuando se utilizan servidores situados fuera del territorio nacional. Lo propio debe observarse respecto de los encargados de datos que, a su vez, podrían utilizar servicios que requieren del envío de los datos personales a territorios extranjeros, en cuyo caso, nuevamente es la organización responsable la que debe garantizar el cumplimiento de todos los requisitos, responder ante el titular y la Agencia.

La salida de los datos desde la custodia de un país hacia otro implica su inserción en un entorno normativo distinto, que puede corresponder a un país con estándares de seguridad menos robustos, donde las autoridades locales cuenten con mayores facultades para acceder a la información o donde exista una mayor exposición a ciberataques que puedan comprometerla.

Asimismo, la trazabilidad de los datos se dificulta y, tanto el cumplimiento de las obligaciones del responsable como el ejercicio de los derechos por parte del titular, pueden tornarse significativamente más complejos. Todo ello puede afectar la capacidad del responsable para garantizar adecuadamente la seguridad de los datos y para gestionar de manera oportuna los derechos que el titular pudiera ejercer, lo que puede significar exponerse a multas.

Para enviar datos personales al extranjero será necesario determinar, en primer lugar, si el país de destino es considerado "adecuado" o seguro. Una vez en funcionamiento, corresponderá a la Agencia publicar el listado de países que, a su juicio, cuentan con un estándar de protección adecuado para Chile.

Para transferir datos a países calificados como seguros, no será necesario adoptar medidas adicionales. En cambio, si el país de destino no se estima adecuado, será obligatorio implementar salvaguardas complementarias, como la celebración de cláusulas contractuales, la adopción de modelos de cumplimiento u otras medidas equivalentes.

Se debe conocer qué se tiene, dónde se tiene, con qué justificación y para qué.

Seguridad y privacidad por diseño y por defecto

Aunque aún queda mucho por abordar, este artículo no puede terminar sin mencionar la seguridad y privacidad por diseño y por defecto, en otras palabras, el rol precavido y proactivo que deberá integrarse a todas las nuevas iniciativas y desarrollos que supongan una modificación relevante a la forma en que la organización trata datos personales o la inclusión de nuevos tratamientos, ya sea por una ampliación de los datos tratados, por la migración a procesamientos masivos o automatizados, o por la necesidad de perfeccionar o desarrollar nuevas soluciones de software para el tratamiento de la información. Este tipo de iniciativas deberán someterse desde el comienzo a revisiones respecto de la necesidad y minimización de los datos, apuntando a la utilización de la menor cantidad posible de datos personales que permita el cumplimiento del objetivo; el análisis de riesgos desde la perspectiva de la seguridad de la información, pero también respecto del impacto del tratamiento sobre las personas; la determinación del ciclo de vida completo de los datos, estableciendo roles y responsabilidades a lo largo de todo su procesamiento; la forma de transparentar al público que este tratamiento se está llevando a cabo, su lógica y fines; los procesos internos para armonizar las necesidades de la organización con el ejercicio de derechos de los titulares; la eventual participación de terceros; el análisis de la seguridad y proporcionalidad de las tecnologías utilizadas, la aplicación de evaluaciones de impacto en los casos en que la ley lo exige y la revisión continua de los proyectos.

Conclusión

En definitiva, la implementación de un nuevo marco normativo de protección de datos personales constituye un desafío interdisciplinario que exige una estrecha coordinación y comunicación entre las áreas jurídicas y técnicas. La incorporación de la seguridad y la privacidad desde el diseño, así como la capacidad de gestionar solicitudes, instrucciones e incidentes, requieren una visión integral y coherente, clave para fortalecer la confianza y la resiliencia de las organizaciones y su convivencia con la ciudadanía. Esta compleja transición nos invita a dialogar, modernizar procesos y consolidar una cultura de cuidado de los datos, de la que todos somos parte. B

Los desafíos para instituciones públicas con la entrada en vigor de la nueva Ley de Protección de Datos Personales

Verónica Achá Álvarez

Master of Science in Public Policy and Management por la Carnegie Mellon University. Jefa de División de Información Social, en el Ministerio de Desarrollo Social y Familia.

@veronicaacha
linkedin.com/in/vacha/

Resumen / Durante años, la protección de los datos personales en el sector público ha estado bajo la supervisión del Consejo para la Transparencia, que instaló criterios y estándares relevantes. Ahora, los nuevos requisitos que entrarán en vigor en diciembre de 2026 plantean desafíos significativos, especialmente considerando que la implementación de la Ley de Transformación Digital también se realizó de forma gradual, segmentando a los órganos del Estado y otorgando plazos diferenciados según sus capacidades.

En este contexto, las instituciones que ya han avanzado en proyectos de gobernanza de datos estarán mejor preparadas para cumplir con las nuevas obligaciones, a diferencia de aquellas que aún no han ordenado sus prácticas de tratamiento de datos. Contar con normativas internas facilita la transición, pero para quienes no han iniciado este camino el desafío será complejo.

Advertencia necesaria

Inicio este artículo declarando que soy hija de Beauchef. Mi formación de base es la ingeniería civil en computación y, si bien he trabajado los últimos 13 años en el Ministerio de Desarrollo Social y Familia, administrando tal vez uno de los registros de datos personales y sensibles más completos del sector público sobre la población que vive en Chile, mi acercamiento ha sido desde la ingeniería. Sin embargo, escribo desde los años de experiencia que tengo trabajando con leyes y su aplicación al tratamiento de los datos personales, camino que he recorrido en un trabajo colaborativo con importantes profesionales del mundo jurídico, tanto en el Ministerio como fuera de él, así como desde el empeño y desafío personal de formarme en la materia, para realizar un trabajo profesional.

No partimos de cero

Por años ha existido en los círculos dedicados al trabajo con datos personales, la convicción de la necesidad de una actualización urgente a la normativa vigente en el país. Y si bien en 2018 se incorporó a la Constitución el derecho a la protección de los datos personales como parte del catálogo de derechos fundamentales [1], todavía era insuficiente frente a una ley anclada a la realidad de 1999, en cuanto a la protección a la vida privada, sin la posibilidad real de ejercer los derechos de las personas frente a abusos sobre ellos.

Sin embargo, desde el punto de vista de la sociedad, las instituciones públicas no estaban en igual nivel de libertad para sus acciones respecto del tratamiento de estos datos, porque como regla general, el derecho administrativo regula la organización y funcionamiento de la administración pública, estableciendo normas para su actuación en el día a día, y principios sobre sus actuaciones, ponen cota y control a ellas.

Sin ir más lejos, como ejemplos, por principio de legalidad, la actuación de los órganos públicos sólo puede realizarse conforme a la ley, y de proporcionalidad, la acción estatal debe ser proporcional a los fines que busca alcanzar.

Así, la ley N° 19.628 vigente [2], contiene un apartado breve donde se fijan parámetros para el tratamiento de datos por parte de los organismos públicos.

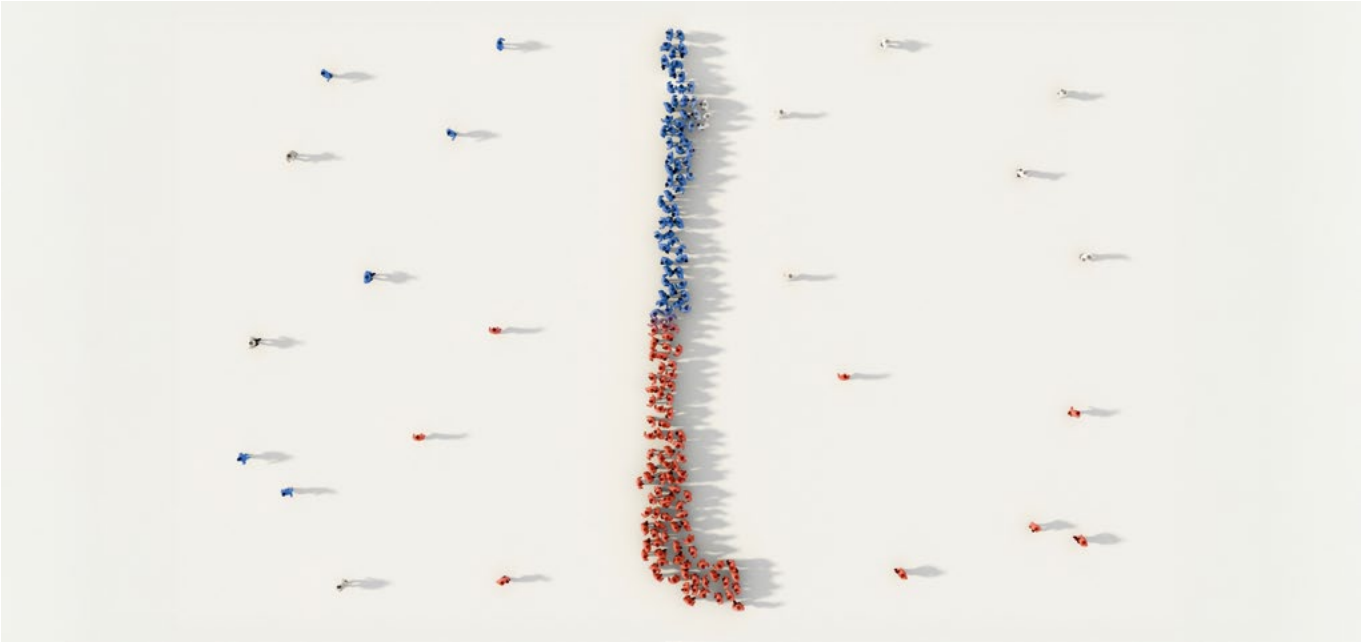
Como segundo elemento de control, no existente para el sector privado, en la ley N° 20.285 se estableció que el Consejo para la Transparencia (CPLT) tendría entre sus funciones y atribuciones "[v]elar por el adecuado cumplimiento de la ley Nº 19.628, de protección de datos de carácter personal, por parte de los órganos de la Administración del Estado" [3]. De tal forma, el CPLT ha cumplido por años un rol semejante al que se espera cumpla la Agencia de Protección de Datos, definiendo lineamientos y estándares, sólo que ahora para todos los sujetos obligados y ya no sólo para las instituciones públicas.

Por tanto, si bien la entrada en vigencia de un nuevo y más alto estándar de control sobre el tratamiento de los datos personales no resulta indiferente, no parte desde cero en cuanto a la necesidad de contar con mecanismos de gestión y control. En lo que sí somos todos los trabajadores de los datos más conscientes del cambio, es que la creación de una Agencia de Protección de Datos, con potestad fiscalizadora para hacer cumplir las disposiciones de la ley, nos lleva a un estándar que hasta ahora no habíamos conocido.

Pero qué tan preparados estamos

Tal vez una de las primeras distinciones que sugiero tener, para entender el impacto de la entrada en vigor de la ley sobre protección de los datos personales, esté dada por la gradualidad establecida para la implementación de la ley de Transformación Digital [4]. En ese caso, se estableció una hoja de ruta —que en los hechos fue ajustándose en el tiempo por el peso de la realidad— dividiendo a los Órganos de la Administración del Estado en tres grupos [5]:

a. *Grupo A.* Conformado por los ministerios; los servicios públicos creados para el cumplimiento de la función administrativa, que no se encuentren en los grupos B y C; la Contraloría General de la República; las Fuerzas Armadas y las Fuerzas de Orden y Seguridad Pública; y las delegaciones presidenciales regionales y provinciales.
b. *Grupo B.* Compuesto por los gobiernos regionales; y un primer conjunto de municipios.
c. *Grupo C.* Integrado por los municipios restantes, no incluidos en el grupo B.



El Consejo para la Transparencia ha cumplido por años un rol semejante al que se espera [ahora] cumpla la Agencia de Protección de Datos.

Esta segmentación y distribución temporal para su implementación tomó en consideración las importantes diferencias organizacionales y de capacidades existentes en el Estado, estableciendo los plazos más cortos y las obligaciones más exigentes para el primer grupo. Como tal, resulta una buena aproximación para avizorar el desafío que significa la implementación de los cambios legales respecto del tratamiento de los datos personales.

Aunque debiera resultar evidente, el tratamiento de datos personales es algo que ocurre en todas las instituciones del Estado. En el más simple y general de los casos, a lo menos las instituciones deben administrar la información de las y los funcionarios que trabajan en ella, en todas las calidades contractuales, con todos los datos de su trayectoria de trabajo y los procesos diarios, como la marca de la asistencia, los mensuales, como el pago de las remuneraciones, y los anuales, como las evaluaciones de desempeño, entre otras. Así mismo, les corresponde administrar los datos de las personas que postulan para acceder a cargos, junto a toda la documentación personal que incorporan en los procesos. Por tanto, si debemos definir, por ejemplo, a quiénes capacitar sobre las nuevas obligaciones, esto debe ocurrir a todo lo largo y ancho de toda la institución, en todos los órganos del Estado.

En cuanto a las instituciones que tienen responsabilidad sobre bancos de datos personales, ellas traen consigo las definiciones y mandatos que el CPLT fue entregando a lo largo de los años, respecto de cómo enfrentar requerimientos a la luz de peticiones por transparencia activa. Y en esa línea, los criterios también fueron ajustándose y cambiando con el tiempo. Ese aprendizaje resulta importante a la hora de la implementación de un estándar más exigente, que conlleva una entidad fiscalizadora y capacidad de persecución más concreta que la disponible hasta ahora.

Otro elemento que permite estimar el nivel de preparación de los servicios públicos frente a las obligaciones de la ley de protección de datos personales es la autoevaluación utilizando el marco de referencia de gestión de datos para los órganos de la Administración del Estado, que forma parte del Sistema de Transformación Digital administrado por la Secretaría de Gobierno Digital. Este instrumento, desarrollado específicamente para el sector público chileno, se basa en el modelo DAMA —un marco global de buenas prácticas para la gestión de datos elaborado por la Data Management Association— al que se incorporaron elementos relevantes para la realidad nacional. Define las áreas de práctica básicas que se espera que todas las instituciones alcancen en un plazo de tres años, entre 2026 y 2028.

Con su aplicación, cada servicio público cuenta con una autoevaluación de este modelo simplificado de gobierno de

datos que le permite identificar brechas para luego elaborar un plan de trabajo trianual, comprometido desde su máxima autoridad, para avanzar en aquellas prácticas que sean especialmente relevantes para su institución y en las que existan brechas por cubrir. Y, aunque este instrumento funciona sólo como un *proxy* para lo que será una gestión adecuada de datos y para las exigencias que impondrá la ley, la gobernanza de datos constituye el punto de partida indispensable para lograr un control apropiado de los datos personales y su protección.

Experiencias compartidas

Los años de tramitación de esta ley dieron oportunidad para conocer cuáles eran las líneas principales que se estaban trabajando, así como algunas preferencias normativas, por ejemplo, que inclinaban la balanza hacia un enfoque más europeo sobre protección de los datos personales. Pequeñas señales como ésta fueron suficientes para que, en instituciones públicas como nuestro ministerio, buscáramos elementos de valor que pudieran aportarnos a un incipiente trabajo en gobierno de datos.

Si bien la Ley sobre Protección de Datos Personales no establece una obligación de realizar gobernanza de datos, la experiencia de tratamiento de grandes bancos de datos enseña que sin un estándar de trabajo es imposible dar cumplimiento a ninguna exigencia. Por ello, el Ministerio de Desarrollo Social y Familia inició, en el año 2018, un trabajo sistemático para gobernar los datos personales y sensibles que le corresponde administrar y, para ello, escogimos el estándar DAMA. La experiencia de este trabajo inicial quedó reflejada en un documento que elaboró la (hoy) Secretaría de Gobierno Digital, con apoyo del Banco Interamericano de Desarrollo [6].

Ya en ese momento, aún sin una nueva normativa, nos desafiamos constantemente a evaluar si éramos capaces de responder a cualquier titular de un dato, qué datos tenemos sobre su persona, con quién lo habíamos compartido, bajo qué amparo legal lo tenemos y lo compartimos, entre otras materias. Tenemos respuestas fáciles y concretas para varias de esas preguntas, pero obtener algunas otras resultaba engorroso o imposible, y sabíamos que debíamos prepararnos para ello. Lo vimos como la necesidad de introducir cambios en formas de trabajo a distintos niveles, que tocan distintas partes de la institución, tanto a nivel funcional o de negocio, como a niveles técnicos y tecnológicos, es decir, era necesario establecer una nueva cultura de trabajo y de conocimiento y de forma de comportamiento institucional respecto del tratamiento de los datos personales.

Como parte de este trabajo, establecimos normativas ministeriales sobre el trabajo con datos personales, para ordenar la forma de trabajo en distintos niveles. Y, si bien no ha estado exento de dificultades, ha permitido entregar señales sobre cómo se trabaja con datos personales, qué significa la proporcionalidad, por qué no todos pueden acceder a los datos que el ministerio trata, qué obligaciones de protección y privacidad tenemos, no sólo como funcionarios públicos sino, específicamente, como funcionarios de este ministerio, entre otras muchas materias, que fuimos dejando reguladas y sobre las que hemos capacitado cada año, una y otra vez, porque los cambios organizacionales y de cultura no ocurren de la noche a la mañana.

Entre las prácticas que instalamos paulatinamente, aprovechando el marco de trabajo DAMA y las prácticas que parecían aplicables de la experiencia europea, estuvo la evaluación de impacto en la privacidad. Para ello, buscamos cursos gratuitos y elaboramos instrumentos propios que nos permitieron comenzar la evaluación de los proyectos tecnológicos ministeriales que usan datos personales, para determinar si en algún punto, significan un riesgo en la privacidad para los titulares de los datos que utilizan. Desde que partimos con esta práctica a la fecha se han evaluado un par de decenas de proyectos ministeriales, tras cuyos resultados se introdujeron ajustes de todo tipo, para hacerlos más seguros desde el punto de vista de la protección de los datos de las personas.

La evaluación de impacto en la privacidad es una de las obligaciones que trae la ley de protección de datos personales, específicamente en su artículo 15 ter y, claramente, nuestro ministerio califica en la obligatoriedad de su realización, por el tipo de tratamiento de datos que realiza.

Volviendo la mirada hacia la Administración del Estado, muchas instituciones que son parte del Grupo A antes mencionado, y que administran bancos de datos, iniciaron proyectos de implementación de gobernanza de datos más o menos a la par de nuestro ministerio. Con muchos de ellos compartimos e intercambiamos experiencias durante estos años, aprendiendo de la forma en que realizamos el trabajo con los datos y del cómo establecimos un modelo de gobierno de datos, para avanzar en el cambio cultural y organizacional que significa lograr esa gobernanza.

De estos diálogos, evalúo que para las instituciones que iniciaron el trabajo de gobernar sus datos con anticipación, será más fácil dar cumplimiento a las obligaciones que se establecen en la normativa que entrará en vigor a contar de diciembre de 2026, en comparación con aquellas que no han iniciado ese camino. Porque el tratamiento de los datos gobernados es algo que se obtiene de manera progresiva a nivel institucional. No pasa por una instrucción o un lineamiento. Se requiere lograr un cambio de cultura de trabajo en torno a los datos personales.

Sin ir más lejos, instituciones como la nuestra, donde hemos trabajado con un foco en datos qué podríamos llamar "de negocio" y no tanto en los datos operacionales, tenemos que trabajar para involucrar a nuevos usuarios que, hasta hoy, no había sido necesario incluir tan vigorosamente, como los equipos de Desarrollo de las Personas y, en general, las personas de Administración y Finanzas, que administran los datos personales de las y los funcionarios y trabajadores de las instituciones.

Nos desafiamos constantemente a evaluar si éramos capaces de responder a cualquier titular de un dato qué datos tenemos sobre su persona, con quién lo habíamos compartido y bajo qué amparo legal lo tenemos y lo compartimos.

Conclusión

La entrada en vigor de la nueva ley de protección de datos personales representa, sin duda, un cambio profundo para las instituciones públicas, pero no un salto al vacío. El sector público arriba a este nuevo estándar con una trayectoria previa: años de regulaciones administrativas, principios de actuación que han marcado límites claros, y el rol fiscalizador que históricamente ejerció el Consejo para la Transparencia. Todo ello configura un punto de partida que, si bien es insuficiente frente a las exigencias actuales, dista de ser inexistente.

Sin embargo, el desafío real no está sólo en la actualización normativa, sino en la capacidad de cada institución para transformar esa base en un modelo maduro de gestión y protección de datos. La experiencia acumulada por instituciones que administran bancos de información y la adopción de marcos de trabajo como DAMA han demostrado que la única forma de responder adecuadamente a las nuevas obligaciones es mediante un cambio cultural sostenido y transversal.

En esa línea, herramientas como el marco de gestión de datos del Sistema de Transformación Digital permiten dimensionar brechas y orientar un trabajo planificado. Si bien es una autoevaluación y, como tal, tiene muchas limitaciones, hoy es el mejor *proxy* al estado de preparación de los servicios públicos respecto de la entrada en vigor de la Ley de Protección de Datos Personales.

Las instituciones que han iniciado el camino del gobierno de datos con anticipación —instalando prácticas, ajustando procesos, definiendo normativas internas y preparando equipos— enfrentarán este nuevo escenario con mayor madurez y capacidad de adaptación. Para aquellas que aún no han comenzado, el desafío será enorme, no sólo porque el tiempo apremia, sino porque la gobernanza de datos no se decreta: se construye día a día, en la operación cotidiana, en la toma de decisiones, en la convicción de que la protección de los datos personales es parte esencial del quehacer público, y eso toca demasiados puntos de la administración. Se requiere un proyecto para lograr gobernarlos, porque requiere un proceso de maduración institucional. **B**

Referencias

[1] Decreto 100 fija el texto refundido, coordinado y sistematizado de la Constitución Política de la República de Chile, Capítulo III De los derechos y deberes constitucionales, artículo 19, número 4°, https://bcn.cl/2eph2.

[2] Ley N° 19.628 sobre protección de la vida privada, Del tratamiento de datos por los organismos públicos, https://bcn.cl/2eqfn.

[3] Ley N° 20.285 sobre acceso a la información pública, artículo 33, letra m), https://bcn.cl/25bya.

[4] Ley N° 21.180 transformación digital del Estado, https://bcn.cl/2eqqx.

[5] DFL 1 establece normas de aplicación del artículo 1 de la ley N° 21.180, de transformación digital del Estado, respecto de los procedimiento administrativos regulados en leyes especiales que se expresan a través de medios electrónicos y determina la gradualidad para la aplicación de la misma ley, a los Órganos de la Administración del Estado que indica y materiales que les resultan aplicables, artículo 5°, https://bcn.cl/yW4svd.

[6] Gobierno de datos en Ministerio de Desarrollo Social y Familia, https://digital.gob.cl/biblioteca/estudios/gobierno-de-datos-en-ministerio-de-desarrollo-social-y-familia/.

Dime qué dicen los datos y no podré decirte quién eres

Cómo publicar información sensible sin comprometer la privacidad de las personas

Matías Toro

Doctor en Ciencias Mención Computación por la Universidad de Chile. Profesor Asistente del Departamento de Ciencias de la Computación de la misma Universidad e Investigador Joven del Instituto Milenio Fundamento de los Datos. Líneas de investigación: lenguajes de programación, sistemas de tipos y privacidad diferencial.

mtoro@dcc.uchile.cl

Resumen / Vivimos rodeados de datos sensibles, y publicarlos sin cuidado puede exponer información sensible, incluso después de procesos de "anonimización". Técnicas clásicas como el *k*-anonimato y la ℓ-diversidad ayudan a ocultar identidades agrupando registros, pero son vulnerables cuando existen datos externos capaces de recombinar o inferir información sensible.

La privacidad diferencial ofrece un enfoque más robusto: en lugar de proteger la tabla publicada, protege el mecanismo que genera los datos, garantizando que el resultado cambie muy poco si una persona participa o no. Esto se logra añadiendo ruido cuidadosamente calibrado, lo que permite mantener la utilidad estadística sin revelar información individual. Hoy es el estándar usado por Google, Apple, Meta y el Censo de Estados Unidos, y será clave para cumplir la nueva Ley de Protección de Datos en Chile.

Introducción

Vivimos rodeados de datos. Cada vez que navegamos por Internet, usamos una tarjeta de transporte, pagamos con el teléfono, vamos al médico o simplemente escuchamos música, generamos información. Estos datos pueden incluir aspectos profundamente personales: nuestra salud, patrones de movilidad, historial financiero, hábitos de citas, compras, gustos artísticos, registros académicos y mucho más. Y, como si fuera poco, los generamos en volúmenes crecientes día a día.

Toda esta información abre un mundo de oportunidades. Permite personalizar servicios, automatizar tareas mediante modelos computacionales, orientar políticas públicas con evidencia, mejorar la transparencia del Estado, e incluso entrenar modelos avanzados de inteligencia artificial. Sin embargo, junto con estas posibilidades aparece un riesgo ineludible: muchos de esos datos contienen información extremadamente sensible. Su uso inadecuado —o su publicación sin suficiente protección— puede exponer facetas íntimas de la vida de una persona.

Este riesgo no es teórico. Hay casos emblemáticos que lo evidencian con crudeza. En 2007, Netflix lanzó una competencia abierta para mejorar su sistema de recomendación y publicó un conjunto de datos "anonimizado". Bastó cruzarlo con información pública de IMDb para que participantes de la competencia lograran reidentificar a numerosas personas, reconstruyendo parte de su historial de visualización.

Un caso aún más icónico ocurrió en 1997, cuando el gobernador de Massachusetts publicó registros médicos de funcionarios públicos tras un proceso de anonimización. Dos días después, Latanya Sweeney —entonces estudiante de doctorado en el MIT— le envió al gobernador una carta con *sus propios* registros médicos. ¿Cómo lo hizo? Cruzó los datos anonimizados con el padrón electoral usando solo tres atributos: código postal, fecha de nacimiento y sexo. La anonimización falló al no considerar información auxiliar disponible para un atacante.

Estos ejemplos ilustran un punto crucial: *el problema no está solo en los datos publicados, sino en los datos externos que podrían combinarse con ellos.*

Estos desafíos adquieren hoy una urgencia especial. La nueva Ley de Protección de Datos Personales, que entrará en vigencia en diciembre de 2026, establece un marco mucho más estricto para el manejo y publicación de información, obligando a organismos públicos y privados a asegurar que cualquier dato liberado —incluso después de procesos de anonimización— no permita identificar a una persona. La ley introduce sanciones importantes y exige adoptar estándares modernos de privacidad. En este escenario, comprender las técnicas que podemos usar para alcanzar ese objetivo se vuelve clave no sólo para investigadores y desarrolladores, sino para cualquier institución que aspire a liberar datos de forma responsable y legal.

Entonces, ¿cómo podemos publicar —o incluso resumir— datos sin revelar información sensible sobre las personas? En la práctica, existen dos grandes enfoques de privacidad:

- **El modelo basado en ataques específicos.** Supone un conjunto definido de amenazas —como los cruces de bases de datos mencionados— y busca defenderse de ellos. Aquí surgen técnicas clásicas de anonimización, como el *k*-anonimato y la *l*-diversidad.

- **El modelo basado en la desinformación controlada.** Su principio es distinto: una publicación es privada si, al observarla, un atacante obtiene *muy poca* información adicional respecto de lo que ya sabía. Este es el fundamento de la *privacidad diferencial*, hoy estándar en proyectos de Google, Apple, Meta y el Censo de Estados Unidos.

Es importante notar que, en marcos legales como la ley de protección de datos, ambos enfoques caen bajo el paraguas de "anonimización", pues buscan impedir que alguien pueda vincular datos sensibles con una persona específica. Pero conceptualmente funcionan de formas muy distintas.

En las siguientes secciones exploraremos estas técnicas, cómo protegen (o no) frente a distintos tipos de ataques, y porqué la privacidad diferencial ha emergido como un cambio de paradigma en la publicación segura de datos.

Anonimización clásica

La anonimización tradicional clasifica los atributos de un conjunto de datos en cuatro categorías:

1. **Identificadores explícitos:** permiten identificar directamente a una persona (por ejemplo, el RUT).

2. **Cuasi-identificadores:** no identifican por sí solos, pero sí cuando se combinan con otros datos. Este fue justamente el caso del ZIP, fecha de nacimiento y género utilizado por Latanya Sweeney para reidentificar al gobernador de Massachusetts.

3. **Atributos sensibles:** contienen información particularmente delicada, como condiciones médicas, historial financiero, antecedentes criminales o nivel de ingresos.

4. **Atributos no sensibles:** todo atributo que no calza en las categorías anteriores.

Podemos analizar los distintos tipos de atributos considerando la Tabla 1, la cual presenta un conjunto de datos ilustrativos sobre personas y su estado COVID-19.

En este ejemplo, *Nombre* es un identificador explícito; *Sexo* y *Dirección* son cuasi-identificadores; y *COVID* es un atributo sensible.

Eliminar nombres no basta

Cuando se publican datos, la técnica más intuitiva consiste en eliminar los identificadores explícitos. Por ejemplo, el conjunto de datos mostrado en la Tabla 2 presenta una versión anonimizada únicamente mediante la supresión del atributo *Nombre*.

Sin embargo, como ya vimos en los casos emblemáticos, *esto no es suficiente* para proteger la privacidad. En particular, no protege frente a los dos tipos de ataques más comunes:

- **Ataques de asociación de registros:** reasocian una fila específica con una persona.

- **Ataques de asociación de atributos:** intentan inferir un atributo sensible sobre una persona, aunque no sepamos exactamente qué fila es.

En el ejemplo, si sabemos que Pedro está en el dataset, que es hombre y vive en *Los Tilos 61*, podemos concluir que le corresponde la fila 1 y, por lo tanto, que tuvo COVID. Esto es un ataque de asociación de registros.

El problema no está sólo en los datos publicados, sino en los datos externos que podrían combinarse con ellos.

N°	Nombre	Sexo	Dirección	COVID
1	Pedro	M	Los Tilos 61	+
2	José	M	Los Alerces 74	+
3	Karla	F	Pasaje Aurora 331	+
4	María	F	Gran Avenida 8585	-

Tabla 1 / Registro de personas y estado de resultado COVID-19.

N°	Sexo	Dirección	COVID
1	M	Los Tilos 61	+
2	M	Los Alerces 74	+
3	F	Pasaje Aurora 331	+
4	F	Gran Avenida 8585	-

Tabla 2 / Conjunto de datos con identificadores explícitos removidos.

k-Anonimato: protegerse de la reidentificación por filas

Para mitigar este tipo de ataques, Samarati y Sweeney introdujeron la noción de *k-anonimato* [1]. Se trata de una propiedad matemática que, (in)formalmente, se puede describir así:

> Una tabla T satisface *k-anonimato* si, para cada combinación de valores de los cuasi-identificadores, existen al menos *k* filas en T que comparten exactamente esa misma combinación de valores.

Intuitivamente, si una tabla satisface *k*-anonimato con $k \geq 2$, entonces un atacante que conoce los cuasi-identificadores de un individuo (por ejemplo, sexo y dirección) no puede saber cuál de esas *k* filas corresponde a la persona. Mientras mayor sea *k*, más fuerte es la protección: el atacante queda "perdido" entre más candidatos posibles. En la tabla original, el grupo de

hombres que viven en *Los Tilos 61* aparece sólo una vez, por lo que la tabla es *1*-anónima: no ofrece anonimato efectivo.

¿Qué podemos hacer para lograr k ≥ 2?

Existen varias técnicas para transformar una tabla y hacer que satisfaga *k*-anonimato, entre ellas:

- Generalización de datos
  (abstraer valores, perdiendo precisión),
- Supresión de registros
  (eliminar filas completas),
- Supresión de celdas
  (ocultar valores puntuales),
- Anatomización,
- Permutación,
- Perturbación aleatoria, entre otras.

La más común y fácil de explicar es la *generalización*: en vez de publicar el valor exacto, se publica una versión más abstracta. Por ejemplo, podemos reemplazar la dirección exacta por la comuna, tal como se muestra en la Tabla 3.

Tras la generalización, cada combinación de cuasi-identificadores *(Sexo, Comuna)* aparece al menos dos veces. La tabla es *2*-anónima. Un atacante que sabe que Pedro es hombre y vive en Valdivia sólo puede acotar que es o la fila 1 o la 2, pero no distinguir entre ambas.

ℓ-Diversidad: evitar asociar personas a valores sensibles

El *k*-anonimato, sin embargo, no protege contra los ataques de asociación de atributos. En nuestro ejemplo, si sabemos que Pedro está en el dataset, que es hombre y vive en Valdivia, basta mirar la tabla *2*-anónima: todas las filas del grupo *(M, Valdivia)* tienen COVID. No sabemos exactamente cuál fila es Pedro, pero sí podemos concluir su atributo sensible.

Para mitigar este tipo de ataque, Machanavajjhala et al. [2] introdujeron la noción de *ℓ-diversidad*. De nuevo, se trata de una propiedad matemática, que se puede describir así:

> Una tabla satisface *ℓ-diversidad* si, para cada grupo de cuasi-identificadores, existen al menos ℓ valores sensibles distintos dentro de ese grupo.

Si una tabla satisface ℓ-diversidad para ℓ≥2, entonces, incluso si el atacante identifica el grupo al que pertenece una persona, *no puede estar seguro del valor sensible,* porque hay al menos dos posibilidades distintas. Al igual que con *k*-anonimato, valores mayores de ℓ entregan garantías de privacidad más fuertes.

N°	Sexo	Dirección	COVID
1	M	Valdivia	+
2	M	Valdivia	+
3	F	La Cisterna	+
4	F	La Cisterna	-

Tabla 3 / Conjunto de datos con direcciones generalizadas a nivel de comuna.

N°	Sexo	Dirección	COVID
1	-	Chile	+
2	-	Chile	+
3	-	Chile	+
4	-	Chile	-

Tabla 4 / Conjunto de datos con cuasi-identificadores altamente generalizados

Es interesante notar que ℓ-diversidad implica *k*-anonimato: si en un grupo hay al menos ℓ valores sensibles distintos, necesariamente ese grupo tiene al menos ℓ registros, por lo que también es al menos ℓ-anónimo.

En la tabla anterior, el grupo *(M, Valdivia)* tiene un solo valor sensible ("+"), por lo que la tabla sólo satisface *1*-diversidad, que es débil: permite ataques de asociación de atributos.

El costo de exigir más privacidad

Intentemos ahora lograr *2*-diversidad sólo mediante generalización. Una opción extrema sería generalizar tanto los cuasi-identificadores que prácticamente los hacemos desaparecer. Esto se observa en la Tabla 4, donde Sexo y Comuna han sido reemplazados por categorías tan amplias que se vuelven prácticamente inútiles.

En este punto, hemos perdido casi toda la estructura: ya no distinguimos ni sexo ni comuna. Es como si hubiéramos eliminado esas columnas. Esto ilustra un punto importante:

Figura 1 / Ejemplo ilustrativo de una imagen original y su versión con ruido pixel a pixel.

> *A medida que aumentamos k o ℓ para obtener más privacidad, los datos tienden a perder precisión y, con ello, utilidad.*

La anonimización clásica está atrapada en una tensión inevitable: no existe "privacidad perfecta" con "utilidad perfecta" al mismo tiempo. Proteger más implica, en general, publicar datos más agregados, menos detallados y menos útiles para ciertos análisis.

La gran limitación de la anonimización clásica

Para empeorar las cosas, incluso si logramos tablas que satisfacen *k*-anonimato y ℓ-diversidad, estas técnicas siguen teniendo una debilidad estructural: dependen fuertemente de la información auxiliar que podría tener un atacante.

Si alguien conoce suficiente contexto sobre una persona —por ejemplo, sabe que no pertenece a ciertos valores sensibles posibles, o tiene acceso a bases de datos externas muy informativas—, aún podría inferir su atributo sensible, pese a las transformaciones.

Este es el gran problema de la anonimización clásica: sus garantías de privacidad son relativas a lo que suponemos que el atacante *no* sabe.

Para mitigar esta dependencia de la información auxiliar, se ha desarrollado una técnica más moderna, privacidad diferencial, que ofrece garantías formales de privacidad incluso frente a atacantes con mucha información extra. En la siguiente sección describiremos en qué consiste y cómo se compara con estas técnicas clásicas.

Privacidad diferencial

La *privacidad diferencial* (DP) [3] representa un cambio de paradigma frente a las técnicas clásicas de anonimización como el *k*-anonimato y la ℓ-diversidad. A diferencia de ellas, la privacidad diferencial no es una propiedad de los datos, sino una propiedad formal —matemática— del mecanismo que genera o publica esos datos.

Su idea central es sorprendentemente intuitiva:

> *El resultado de un análisis debe ser prácticamente el mismo haya o no participado una persona en la base de datos.*

Esto calza con una noción muy natural de privacidad: *"Nada malo debería ocurrirme como resultado de participar en un estudio. Si algo malo pasa, habría pasado igual incluso si yo no hubiera participado"*.

Lo más importante es que la privacidad diferencial es independiente de la información auxiliar que pueda poseer un adversario. Esto incluye escenarios extremos, como cuando el atacante conoce toda la base de datos excepto la fila objetivo. En este sentido, la privacidad diferencial previene los ataques clásicos que afectan al *k*-anonimato y a la ℓ-diversidad, y se acerca —según lo que se conoce hoy— a la única forma de anonimización "verdadera" en sentido robusto.

Una definición (ligeramente) más formal

Sea un mecanismo *F* que recibe una base de datos y produce un resultado numérico. Decimos que *F* es ε-*diferencialmente*

privado (ε-DP) si, al aplicarlo a dos bases de datos idénticas excepto por un solo individuo, los resultados que entrega son indistinguibles. Esta indistinguibilidad se logra mediante ruido agregado de forma cuidadosamente calibrada. La belleza de DP está en su universalidad: no restringe qué sabe el adversario, no impone supuestos sobre el mundo externo y no depende de clasificar columnas como "cuasi-identificadores" o "sensibles". Es una propiedad matemática pura del mecanismo.

Una intuición visual: ruido a nivel de píxeles

Para construir intuición, imaginemos que queremos publicar una imagen. En la Figura 1, la imagen original está a la izquierda, mientras que a la derecha, mostramos la misma imagen, pero con ruido añadido pixel a pixel.

Si hacemos zoom sobre un pixel particular, el ruido hace que sea imposible saber si el color original era más claro o más oscuro. Este fenómeno recibe el nombre de *denegación plausible*: el pixel puede "negar" haber tenido su valor real porque el ruido lo enmascara.

Pero si miramos la imagen completa, podemos seguir distinguiendo su estructura global: contornos, colores predominantes, patrones. Esto es exactamente lo que permite la privacidad diferencial: perder precisión en lo individual, manteniendo utilidad estadística a nivel agregado.

Si añadimos muy poco ruido, la privacidad se pierde; si añadimos demasiado, como en la Figura 2, la utilidad desaparece y obtenemos un manchón irreconocible. El hilo conductor de DP es encontrar el punto intermedio óptimo.

¿Cómo se decide cuánto ruido agregar?

Todo depende de la sensibilidad de la consulta: cuánto puede cambiar el resultado si cambia un solo individuo.

Por ejemplo, para la consulta *f(x) = #(personas con COVID)*, la sensibilidad es 1, porque agregar o quitar a alguien cambia el resultado como máximo en 1.

Con esto, una forma más básica de crear un mecanismo diferencialmente privado es:

$$F(x) = f(x) + Laplace(s/\varepsilon)\,,$$

donde *s* es la sensibilidad de la consulta, ε el *presupuesto de privacidad*, y Laplace es la distribución de ruido utilizada. Así, *F* es un mecanismo ε-DP.

Eliminar nombres no basta [...] para proteger la privacidad.

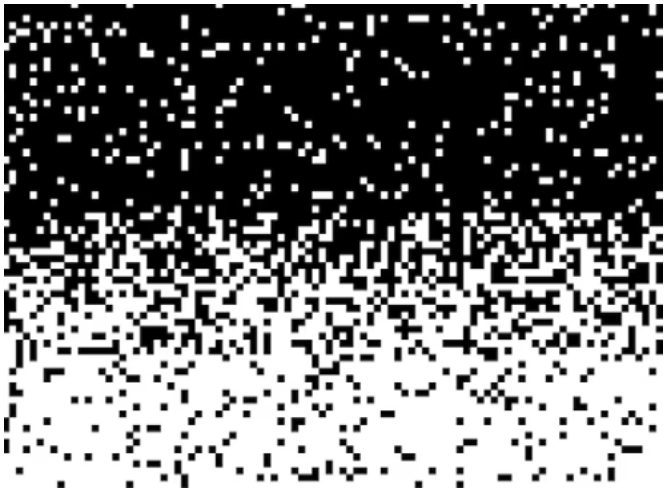


Figura 2 / Ejemplo de imagen con ruido extremo donde se pierde completamente la utilidad.

A mayor *ε*: menos ruido → más utilidad, pero menos privacidad.
A menor *ε*: más ruido → más privacidad, pero menos precisión.

Elegir ε es difícil y requiere experimentación: es uno de los temas más activos en la literatura actual.

Ejemplo numérico

Si el número real de personas con COVID es 14.237, una versión 2-DP podría devolver valores como 14.211, 14.260 o 14.237 (casualmente), valores cercanos, plausibles, y estadísticamente útiles.

Es importante mencionar que aquí usamos la *propiedad de post-procesamiento*: cualquier modificación posterior al resultado (por ejemplo, truncarlo a enteros) no disminuye la garantía ε-DP.

Consultas maliciosas y la importancia del ruido

Consideremos una consulta maliciosa: *g(x) = #(personas llamadas "Alan Brito" con COVID)*. Supongamos que en la base real el valor es 1. Una versión *2*-DP podría dar: 18, 47, –0,68. Estos resultados *no son útiles*, pero está bien: la consulta en sí misma era peligrosa. La privacidad diferencial está diseñada precisamente para impedir que consultas hiperespecíficas o maliciosas revelen secretos individuales.

¿Y si repetimos la consulta muchas veces?

El lector atento notará que si consultamos repetidamente, con ruido distinto cada vez, podríamos aproximar el valor original a través de un histograma, como se ilustra en la Figura 3.

Para evitarlo existe el *presupuesto de privacidad* ε: cada consulta "consume" parte del presupuesto y el sistema debe dividirlo entre las consultas.

Esto se formaliza mediante la propiedad de *composición secuencial*:

Si un mecanismo publica dos resultados, uno ε_1-DP y otro ε_2-DP, entonces el mecanismo completo es $(\varepsilon_1 + \varepsilon_2)$-DP.

Por ejemplo, si queremos publicar cuatro veces una consulta con presupuesto total ε, cada una debe usar $\varepsilon/4$, lo que produce resultados más ruidosos e impide reconstrucciones maliciosas.

Pero aquí aparece una dificultad importante: si hacemos muchas consultas, el ε disponible para cada una se vuelve muy pequeño, y el ruido agregado puede crecer hasta un punto en que la utilidad disminuye drásticamente.

Composición paralela: cuando los datos no se solapan

Afortunadamente, existe otra propiedad fundamental de la privacidad diferencial: la *composición paralela*.

Si múltiples mecanismos ε-DP se aplican a conjuntos disjuntos de individuos, el mecanismo conjunto sigue siendo ε-DP.

Esto tiene una consecuencia práctica muy valiosa: si las consultas se aplican sobre partes distintas de la población, no es necesario pagar el costo de la composición secuencial, es decir, no se "consume" más presupuesto de privacidad. En la práctica, podemos aplicar varias consultas ruidosas "gratis" en términos de ε.

Esto es especialmente útil para publicar histogramas, donde cada celda (o *bin*) corresponde a un grupo distinto de personas. Por ejemplo, si queremos contar casos de COVID separados por sexo y comuna, podemos aplicar ruido laplaciano a cada celda sin dividir ε, porque cada celda involucra individuos diferentes y, por lo tanto, cumple las condiciones de la composición paralela. Por ejemplo, podemos partir de la Tabla 5, que contiene los conteos exactos de casos para cada combinación de sexo y comuna.

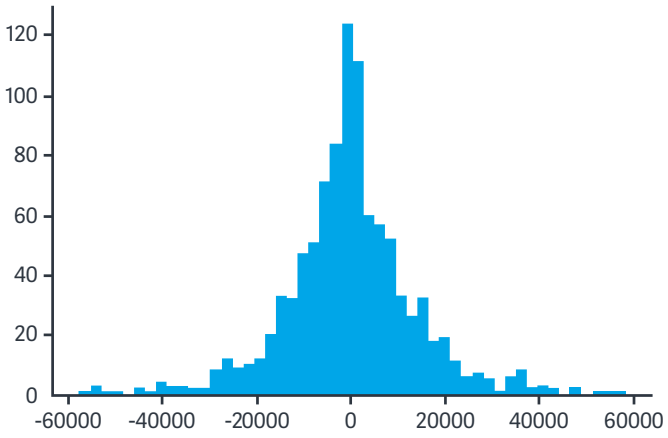


Figura 3 / Histograma obtenido al repetir una consulta con ruido Laplace.

Sexo	Comuna	COVID
M	Valdivia	295
F	Valdivia	743
M	La Cisterna	289
F	La Cisterna	122

Tabla 5 / Conteo real de casos de COVID por sexo y comuna.

Sexo	Comuna	COVID
M	Valdivia	296
F	Valdivia	740
M	La Cisterna	286
F	La Cisterna	126

Tabla 6 / Conteo publicado con ruido laplaciano y redondeo.

Si aplicamos ruido Laplaciano según una distribución *Laplace(½)* a cada celda del histograma, luego, gracias a la propiedad de postprocesamiento de la privacidad diferencial, podemos truncar los valores al entero más cercano sin perder la garantía de privacidad. El resultado final puede observarse en la Tabla 6, que muestra el histograma publicado después de añadir ruido y redondear los valores. En este caso, la utilidad estadística se preserva y la garantía ε-DP permanece totalmente vigente.

Más allá de consultas agregadas: datos sintéticos y modelos de IA

Como la privacidad diferencial es una propiedad de mecanismos, no de tablas, su aplicación va mucho más allá de consultas individuales. Hoy en día se usa para: generar datos sintéticos con ruido estadístico controlado, o para entrenar modelos de aprendizaje automático con privacidad garantizada.

En este último caso, la garantía es potente: si un modelo fue entrenado con tus datos de forma diferencialmente privada, entonces su comportamiento sería prácticamente el mismo aunque tus datos no hubieran estado en el entrenamiento. En otras palabras: Participar o no en el entrenamiento no cambia lo que el modelo puede revelar sobre ti. Esto es extremadamente relevante en la era de los modelos generativos, donde fragmentos de datos de entrenamiento pueden "colarse" en outputs inesperados.

Conclusión: ¿Cómo se comparan la anonimización clásica y la privacidad diferencial?

Las técnicas de anonimización clásica —como el *k*-anonimato y la ℓ-diversidad— buscan reducir el riesgo de reidentificación eliminando identificadores y agrupando datos. Son fáciles de aplicar y ampliamente usadas, pero dependen fuertemente de cómo se seleccionan y transforman los cuasi-identificadores. Además, son vulnerables a la información auxiliar: un adversario bien informado puede, en muchos casos, reconstruir o inferir atributos sensibles incluso cuando la tabla cumple las propiedades requeridas.

La belleza de DP está en su universalidad: no restringe qué sabe el adversario, no impone supuestos sobre el mundo externo y no depende de clasificar columnas.

La privacidad diferencial, en cambio, sigue un enfoque completamente distinto. En vez de proteger los datos publicados, es una propiedad del mecanismo que produce esos datos, garantizando que el resultado sea prácticamente indistinguible haya o no participado un individuo. Su fortaleza clave es que la garantía no depende del conocimiento del adversario, incluso si este conoce todo menos un registro. Por eso se considera la técnica más robusta disponible hoy.

En resumen:

- **Anonimización clásica**: útil, intuitiva, pero frágil ante datos auxiliares.
- **Privacidad diferencial**: formal, robusta y conceptualmente alineada con una noción moderna de privacidad.

La entrada en vigencia de la nueva Ley de Protección de Datos Personales en diciembre de 2026 hará que estas distinciones sean especialmente importantes. La ley exige anonimización robusta y evaluaciones de riesgo considerando posibles ataques de reidentificación. En este escenario, la privacidad diferencial emerge como una herramienta clave para cumplir con las nuevas obligaciones y proteger adecuadamente a las personas al publicar o compartir datos. **B**

Agradecimientos

Gracias a Arturo Kullmer por proporcionar los ejemplos usados aquí para explicar anonimización y privacidad diferencial.

Referencias

[1] Pierangela Samarati y Latanya Sweeney. 1998. *Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression.*

[2] Ashwin Machanavajjhala, Johannes Gehrke, Daniel Kifer, y Muthuramakrishnan Venkitasubramaniam. 2006. *L-diversity: Privacy beyond k-anonymity.* 22nd International Conference on Data Engineering (ICDE'06), 24–24.

[3] Cynthia Dwork. 2006. *Differential Privacy.* En Automata, Languages and Programming, 33rd International Colloquium, ICALP 2006, Proceedings, Part II (LNCS, Vol. 4052). Springer, 1–12.

Datos personales, datos de vida

Patricio Inostroza

Profesor Asistente del Departamento de Ciencias de la Computación de la Universidad de Chile. Docteur en Informatique por la Université Joseph Fourier, Francia, Diplomado en Protección de Datos Personales por la Universidad de Chile e Ingeniero en Computación por la Universidad de Chile. Autor del curso Introducción al Derecho Informático, dictado para la carrera de Ingeniería en Computación, en la Universidad de Chile.

patricio.inostroza@dcc.uchile.cl

Resumen / La Ley 21.719 de Protección de Datos Personales (PDP) de Chile, promulgada en diciembre de 2024, entrará en pleno vigor en diciembre de 2026. Esta ley proactiva devuelve al titular (persona natural) el control sobre sus datos personales, estableciendo que estos derechos son intransferibles e irrenunciables.

La ley se basa en ocho principios fundamentales y seis derechos claves. Define actores específicos: el Responsable (quien responde por el cumplimiento), el Encargado (quien procesa datos), el Delegado (supervisor interno) y la Agencia de Protección de Datos Personales (organismo fiscalizador autónomo).

Todas las instituciones, públicas y privadas, manejan datos personales. El cumplimiento normativo requiere concientización de toda la organización. Si no se logra cumplir con la ley, las sanciones son severas, pudiendo alcanzar los 1.385 millones de pesos.

El 27 de junio de 2025, el diario El País de México informó que el Cartel de Sinaloa hackeó al FBI para asesinar a sus informantes en México: *"...el pirata informático... pudo obtener de su dispositivo el registro de llamadas realizadas y recibidas, así como los datos de geolocalización..."*. Si bien es una noticia sorprendente, cabe señalar que no es la primera vez que los datos personales se usan con fines letales.

Durante la Segunda Guerra Mundial, cuando los nazis invadían un país, enseguida se apoderaban de los registros locales como primer paso para controlar a la población y, en particular, para localizar a los judíos [1].

En mi infancia, el hospital de la región donde nos atendíamos tenía como política que la ficha médica no se entregaba al paciente bajo ninguna circunstancia. Si una persona cambiaba de comuna, se iba sin su ficha, sin la información de su propio historial de salud.

Los ejemplos anteriores muestran que los datos personales no sólo representan un valor económico: su control puede incluso poner en riesgo la vida de las personas.

Cuando el hospital negó el acceso al expediente, generó una dependencia de su "cliente". El paciente no tenía el control de sus propios datos personales. Cuando los nazis tuvieron acceso a los datos locales, pudieron aplicar sus políticas de discriminación y exterminio. Los datos personales les otorgaron un enorme poder sin control. Cuando el cartel de Sinaloa hackeó los datos, puso en evidencia que la seguridad requiere una mirada seria y cuidadosa.

¿Cómo enfrentar estas y otras situaciones donde el uso de los datos personales está en cuestionamiento?

La Ley 21.719 [2] vio la luz el 13 de diciembre de 2024. Conocida como Ley de Protección de Datos Personales (PDP), fue un gran paso para enfrentar las situaciones anteriores. Si bien la ley está en un periodo de transición, este periodo termina en diciembre de 2026. Los dos años que transcurrirán fue el tiempo que la misma ley estableció como suficiente para que las empresas se adapten y alcancen el pleno cumplimiento.

Los datos personales no sólo representan un valor económico: [dan poder y] su control puede incluso poner en riesgo la vida de las personas.

Pero esta ley tiene una sutil diferencia frente a otras leyes que nos gobiernan: es una ley **proactiva**. Pero no nos adelantemos; lo mejor es partir en orden.

El poder de los datos personales

Las situaciones mencionadas al comienzo de este artículo presentan características comunes: los datos personales dan poder. Acumular datos personales es acumular poder; ese poder requiere control, requiere cuidado. Pero ¿quién es el verdadero dueño de los datos personales?, ¿a quién le pertenece ese poder?, ¿quién debe controlarlo?, ¿cómo aseguramos que ese poder no caiga en manos equivocadas? y ¿cómo garantizamos que el cumplimiento sea efectivo?

La Ley 21.719 busca dar respuesta a estas y otras inquietudes.

¿Quién es el titular de los datos personales?

Formalmente, el titular es la persona natural, identificada o identificable, a quien conciernen o se refieren los datos personales.

Esta definición pone de manifiesto que sólo personas naturales pueden ser titulares de datos personales. Luego, una empresa, organización o institución no será titular de datos personales, ya que no es persona natural.

Más aún, la ley retorna al titular el control y la autonomía de su información personal. Además, establece expresamente que los derechos del titular son intransferibles e irrenunciables.

¿Qué es un dato personal?

La respuesta puede parecer obvia, pero no lo es. Se tiende a pensar en el nombre, en el número del documento de identidad de una persona (RUT en Chile) o en la fecha de nacimiento, lo cual es correcto, pero hay sutilezas a cuidar.

El GPS de una camioneta no sería un dato personal, pero si a la camioneta se le asigna un chofer en particular, ahora el dato del GPS se ha transformado en un dato personal, ya que identifica a una persona. Esto último es la clave de todo.

La ley define: *"Dato personal: ...cualquier información vinculada o referida a una persona natural identificada o identificable. Se considerará identificable toda persona cuya identidad pueda determinarse, directa o indirectamente..."*

Por lo tanto, lo que para una empresa es solo un dato industrial, para otra institución puede ser un dato personal. Luego, cada organismo privado o público debe hacer un levantamiento para conocer qué tipo de datos tiene: el primer paso para la clasificación de sus datos.

¿Cómo evitar que la ley se vuelva obsoleta?

A fines de los años ochenta, Chile logró estar conectado a Internet; el mundo comenzó a digitalizarse. En el año 1999 se promulgó la Ley 19.628, Sobre Protección de la Vida Privada. Pero esta ley pronto quedó obsoleta, ya sea porque no cubría todas las situaciones que el mundo digital comenzó a generar, o porque no había un equilibrio entre las sanciones y los beneficios que podía recibir una empresa cuando realizaba un tratamiento de datos personales que no cumplía con la ley. A esto se sumó la falta de una entidad que verificase, de forma proactiva, el cumplimiento de la ley.

Para enfrentar lo señalado, la Ley 21.719 ha establecido con relativa claridad los principios, los derechos y una Agencia para la protección de los datos personales.

¿Para qué sirven los principios en la Ley de PDP?

El espíritu de la Ley 21.719 se ha plasmado en ocho principios. A diferencia de reglas muy detalladas que pueden quedar obsoletas, los principios son más abstractos y perdurables, permitiendo que la ley se mantenga relevante ante cambios tecnológicos y sociales, cambios que son frecuentes en el mundo digital. Los principios establecen los valores o fundamentos rectores que guían cómo debe aplicarse la ley y cómo debe realizarse el tratamiento de datos personales.

Más que detallar qué indica cada principio, el siguiente set de preguntas permite evaluar de forma básica si su empresa o institución ya los cumple[1]:

- **Principio de licitud y lealtad:** ¿Obtuvo los datos personales de forma legal? ¿Hubo alguna triquiñuela para obtener los datos personales?
- **Principio de finalidad:** ¿Ha indicado claramente al titular para qué serán usados los datos personales?
- **Principio de proporcionalidad:** ¿Es realmente necesario contar con cada dato personal que mantiene su organización? ¿Se justifica ese dato para el servicio que se le entrega al titular?
- **Principio de confidencialidad:** ¿Mantiene la confidencialidad siempre, incluso finalizada la relación con el titular de los datos?
- **Principio de seguridad:** ¿Dispone de la seguridad adecuada para evitar fugas o robo de los datos personales? ¿Y cómo está la seguridad de sus proveedores?
- **Principio de transparencia e información:** ¿Ha informado de forma adecuada al titular de lo que hará con los datos personales? ¿Dispone de los medios para que el titular de los datos ejerza sus derechos?
- **Principio de calidad:** ¿Mantiene todos los datos personales actualizados? ¿En todos los sistemas de la empresa?
- **Principio de responsabilidad:** ¿Tiene claro que siempre será el responsable del tratamiento de los datos personales, incluso si delega el tratamiento a un tercero?

1 En rigor, el trabajo de evaluación debe ser realizado por un especialista en Protección de Datos Personales. Este cuestionario sólo tiene como fin mostrar los elementos básicos, por lo que debe ser tomado sólo como un elemento introductorio en el tema.

Al contar con estos ocho principios se tiene una base sólida que permite destrabar situaciones no previstas por la ley. Es un marco que orienta a jueces, autoridades y profesionales al enfrentar nuevas situaciones. Establece cómo debería ser el comportamiento de las instituciones privadas y públicas. Da un marco para enfrentar y adaptarse a nuevas tecnologías. Orienta cómo velar por su cumplimiento. Permite tener una primera aproximación para que profesionales de diversas ramas, como ingenieros, médicos, periodistas y más, puedan evaluar si sus proyectos cumplen con la normativa. Y, sobre todo, protegen al titular de los datos cuando la ley no cubre un caso particular.

¿Cuáles son los derechos en la Ley de PDP?

Al inicio de este artículo presentamos situaciones que mostraban que la relación entre las personas y las organizaciones que recolectaban datos era muy desigual. El titular entregaba su información, pero tenía poca capacidad para saber qué pasaba con ella después.

Seis son los derechos que permiten a las personas ejercer control efectivo sobre sus datos. Gracias a estos derechos, el titular decide qué pasa con su información. Sin estos derechos, el tratamiento quedaría sólo en manos de las empresas u organismos públicos. Ahora las instituciones, bajo solicitud del titular, deben informar qué datos tienen, para qué los usan, si los comparten y cuánto tiempo los conservarán. Con esta información, el titular puede solicitar que se rectifiquen, actualicen o supriman datos con información inexacta, obsoleta o innecesaria. Incluso puede revocar el consentimiento para el tratamiento de sus datos personales, salvo en casos que la misma ley expresamente lo impida.

Frente a la capacidad que tienen las instituciones públicas y privadas para recopilar y procesar grandes volúmenes de datos, los derechos del titular funcionan como un contrapeso que restablece el equilibrio.

- **Derecho de acceso:** El organismo que haga tratamiento de datos personales debe responder al titular si se están tratando sus datos, incluyendo su origen, finalidad y destinatarios.

- **Derecho de rectificación:** Permite al titular solicitar que se modifiquen o complementen datos inexactos, desactualizados o incompletos.

- **Derecho de supresión (o cancelación):** Este derecho permite al titular solicitar que se eliminen los datos cuando ya no sean necesarios, cuando se haya revocado el consentimiento o cuando hayan expirado los plazos de conservación.

La ley retorna al titular el control y la autonomía de su información personal.

- **Derecho de oposición:** Le da derecho al titular de rechazar tratamientos específicos, como decisiones automatizadas o perfiles basados en datos sensibles, salvo excepciones legales.

- **Derecho a la portabilidad:** Si el titular lo solicita, el organismo debe entregar una copia de sus datos personales en formato estructurado. Incluso puede solicitar que se transfieran sus datos a otro responsable (empresa).

- **Derecho de bloqueo:** Finalmente, el titular puede solicitar suspender temporalmente el tratamiento de sus datos personales cuando se impugne su exactitud o se investigue una infracción.

Es necesario hacer presente que no es suficiente que la empresa dé cumplimiento a los derechos solicitados por el titular de los datos personales. En efecto, es importante que además deje registrado todo tratamiento en una plataforma o medio auditable. Toda transacción relacionada con el tratamiento de datos puede ser sometida a cuestionamiento, lo que puede terminar en un litigio y con esto enfrentar un peritaje. Si el registro es digital, el peritaje lo realizaría un informático.

En cada caso deberá demostrar que el requerimiento del titular fue atendido en el plazo estipulado por la ley y que el titular fue informado de la respuesta. Notar que la temporalidad es crucial, es decir, hay que demostrar que la respuesta fue entregada en el plazo que la ley estipula.

¿Cómo la ley protege al titular de los datos personales?

En general, una de las principales falencias de las leyes es que actúan de forma *reactiva*. Sólo cuando una persona u organismo quebranta la ley y es atrapado, se recurre a ella, donde se activa la investigación, la acusación y la defensa. Si lo amerita, recaen las sanciones. Este modelo presenta una grave falencia: a pesar de que puedan existir sanciones, el daño ya está hecho, lo que en muchos casos es irreversible.

La Ley 21.719 ha sido propuesta como una ley *proactiva*, una ley que obliga a cada institución, pública o privada, a velar por el cumplimiento en cada instante. A esto se suman las sanciones, tema que ahondaremos más adelante.

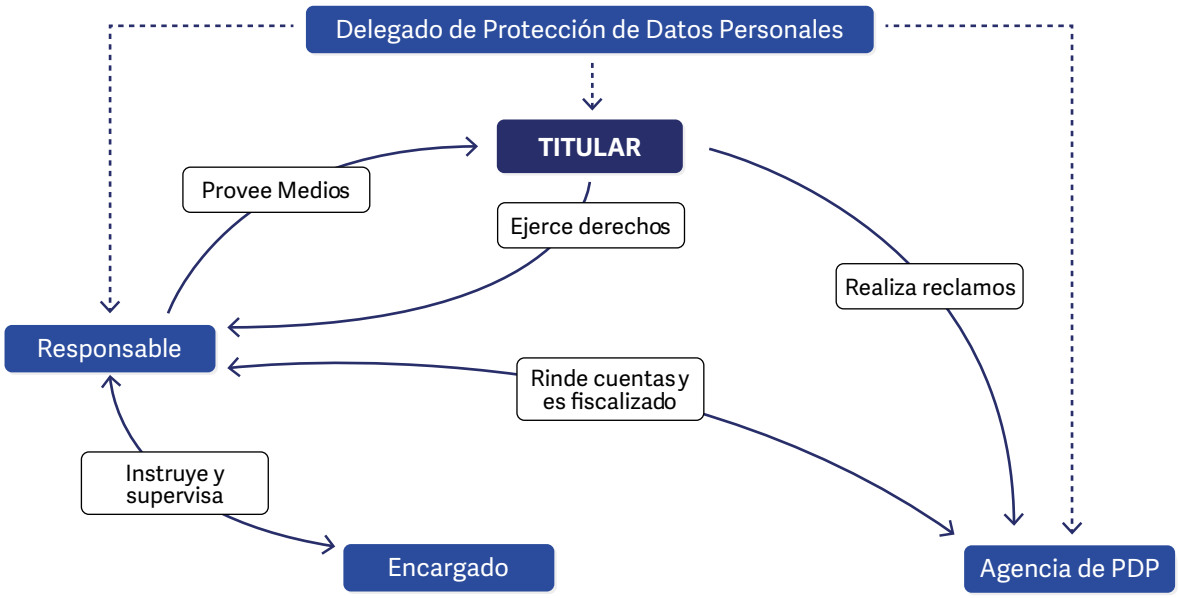


Figura 1 / Esquema simplificado de la interacción de los actores en la Ley 21.719.

¿Cómo es que esta ley tiene actores?

En esencia, la ley crea un sistema donde el *Titular* tiene el poder, el *Responsable* asume la responsabilidad frente a cualquier tratamiento que no cumpla con la ley, el *Encargado* es quien efectúa el tratamiento de datos personales bajo estrictas reglas, donde el *Delegado* asegura una correcta implementación interna. Finalmente, está la *Agencia* de Protección de Datos Personales, encargada de supervisar todo el sistema y de aplicar las sanciones (ver Figura 1).

El titular ya ha sido descrito, por lo que nos centraremos en el resto de los actores.

La empresa o institución debe contar con un *Responsable* y un *Encargado*. El primero es el principal obligado y debe responder por el cumplimiento de la ley. Entre otras actividades, debe velar por el cumplimiento de los derechos del titular, llevar un registro de las operaciones realizadas y, en caso de problemas de seguridad, notificar a la Agencia de Protección de Datos Personales y al titular cuando sea relevante.

A su vez, el *Encargado* es una persona natural o jurídica que trata datos personales. No decide sobre los datos, sólo los procesa siguiendo instrucciones. Cabe señalar que la responsabilidad sigue recayendo sobre el Responsable. Si el Encargado es un servicio tercerizado, será necesario revisar los contratos y velar por que estos cumplan con la ley, es decir, que se preserve la confidencialidad y seguridad de la información, donde además no puedan usar los datos para un fin propio.

Una de las innovaciones más interesantes de la ley corresponde a la creación de la *Agencia* de Protección de Datos Personales. Éste es un organismo público, autónomo y técnico, destinado a supervisar y fiscalizar el cumplimiento de la normativa. Además, media en conflictos entre titulares y responsables, emite instrucciones y guías sobre cómo aplicar la ley e impone severas sanciones (multas), entre otras funciones.

Si bien hay otros actores que escapan del fin de este artículo, hemos dejado para el último al *Delegado* de protección de datos personales. Este es una persona natural designada por el Responsable del tratamiento para actuar como un puente de comunicación entre el Responsable, los Titulares y la Agencia. Su rol es supervisar el cumplimiento interno de la ley y fomentar una cultura de protección de datos dentro de la organización. Su accionar debe ser proactivo y autónomo, teniendo presente que no puede tener conflictos de interés.

Esto último es esencial. Dado que el área TI y el área de recursos humanos realizan tratamiento de datos personales, quedarían inmediatamente excluidos del rol de delegado.

¿Cómo son las sanciones?

Esta ley establece tres tipos de infracciones: leves, graves y gravísimas, donde las multas bordean en su cota superior los 345 millones, 690 millones y 1.385 millones de pesos, respectivamente[2].

Las infracciones leves son de orden administrativo, como incumplir el principio de información y transparencia, no proveer datos para contactar al responsable, dar una respuesta incompleta o fuera de plazo.

En el punto medio, las infracciones graves incluyen problemas como no contar con el consentimiento del titular, tratar datos personales innecesarios, realizar tratamiento de datos personales de niños sin el adecuado resguardo, fallas en la seguridad, vulnerar el secreto o confidencialidad, entre otras.

Y, finalmente, las infracciones gravísimas: tratar datos personales de forma fraudulenta (uso de bases de datos del mercado negro), usar información no veraz, no informar de vulneración en las medidas de seguridad, incluso el incumplir una resolución de la Agencia, son parte de las sanciones más elevadas que puede sufrir una empresa.

¿Y cómo se cumple con la ley?

Cada vez que un empresario, gerente o director de empresa señala que esto es un tema del área de informática, muestra que no ha comprendido lo que implica dar cumplimiento a la ley.

Esta ley afecta a todos los integrantes de la empresa. Si un empleado incumple la ley, si un proveedor que maneja los datos personales sufre una fuga de datos, si recursos humanos recopila los datos de los hijos de los trabajadores para entregar regalos de Navidad pero sin el debido resguardo, la sanción recae en el responsable de la empresa. Notar que en cada ejemplo de este párrafo no intervino el área de TI.

Los principios [que establece la ley] son más abstractos y perdurables, permitiendo que se mantenga relevante ante cambios tecnológicos y sociales.

La palabra clave para dar inicio al cumplimiento de la Ley 21.719 es la *concientización* en el tema, partiendo por la plana mayor y luego a los colaboradores y las empresas o proveedores externos. Sin una adecuada capacitación se genera desconocimiento del problema y aparece la resistencia al cambio.

En paralelo, hay que revisar y actualizar el reglamento interno y los contratos, tanto los de trabajo como los de proveedores. A esto se suma que hay que realizar un levantamiento de los tipos de datos personales que maneja la empresa para verificar que se cumplan los principios. Revisar si los modelos de seguridad interna son los adecuados, si la privacidad se cumple, establecer roles de mínimo acceso a los datos personales. Contar con un registro de cada tratamiento de datos, un registro que soporte una auditoría (peritaje) legal. Muchos de estos temas se cruzan con ISO 27001[5].

La implementación del equivalente de esta ley en Europa y otros países ha generado más de un dolor de cabeza, especialmente a quienes reaccionaron demasiado tarde.

Al momento de publicarse este artículo, restarán pocos meses para cumplir una ley que exige definiciones a nivel de directorio y gerencia, asignación de recursos y una gestión efectiva de los riesgos asociados al tratamiento de datos personales. Esto es un trabajo colaborativo entre abogados, informáticos, asesores, personal de la empresa, proveedores, etc. Poner a todos de acuerdo con el mismo objetivo requiere de planificación. No espere hasta última hora. B

Referencias

[1] Carissa Véliz, "Privacidad es poder: datos, vigilancia y libertad en la era digital". Editor digital: XcUiDi.

[2] Ley 21.719: Regula la Protección y el Tratamiento de los Datos Personales y Crea la Agencia de Protección de Datos Personales. https://www.bcn.cl/leychile/navegar?idNorma=1209272.

[3] La Verdadera y Real Historia de Internet en Chile. https://users.dcc.uchile.cl/~ppoblete/sigloxxi-27Feb96.html.

[4] Ley 19.628 Sobre Protección de la Vida Privada. https://www.bcn.cl/leychile/navegar?idNorma=141599.

[5] ISO/IEC 27001:2022 Information security, cybersecurity and privacy protection — Information security management systems — Requirements. https://www.iso.org/standard/27001.

2 Valores aproximados a la fecha de la publicación de los 5 mil UTM, 10 mil UTM y 20 mil UTM que efectivamente la ley establece.

Tecnologías asistivas para la rehabilitación de niñas y niños portadores de quemaduras:

Construyendo puentes entre computación y fisiatría

Francisco J. Gutiérrez

Doctor en Ciencias mención Computación por la Universidad de Chile. Profesor Asistente del Departamento de Ciencias de la Computación, Universidad de Chile. Líneas de Investigación: interacción humano-computador, computación social.

frgutier@dcc.uchile.cl

María Gabriela Hidalgo

Médico Cirujano, especialista en Medicina Física y Rehabilitación, por la Universidad de Chile.
Jefa de Rehabilitación de la Corporación de Ayuda al Niño Quemado (COANIQUEM). Líneas de Investigación: manejo del dolor, nuevas herramientas terapéuticas en rehabilitación infanto-juvenil.

ghidalgo@coaniquem.org

Resumen / Desde el año 2022, junto a la Corporación de Ayuda al Niño Quemado (COANIQUEM) hemos estado trabajando en la cocreación de prototipos de tecnologías asistivas para apoyar la rehabilitación de niñas y niños portadores de quemaduras, siguiendo un enfoque multidisciplinario y centrado en el paciente. Así, hemos explorado el uso de realidad virtual para mitigar experiencias negativas asociadas al dolor agudo (por ejemplo, durante la curación de heridas), videojuegos que promueven actividad física para ejercitar distintas articulaciones (tales como hombro y codo), y simulaciones interactivas para apoyar ejercicios de terapia conductual cognitiva en la reinserción social de las y los pacientes al completar sus tratamientos. Nuestros resultados preliminares logran dar cuenta de evidencia favorable en la efectividad de los distintos prototipos desarrollados, en particular como complemento a procesos convencionales de rehabilitación, adherencia al tratamiento y potencial de adopción preclínica.

La rehabilitación es un área de la salud que debe ser abordada de manera transdisciplinaria, para lograr sinergias de trabajo en equipo que maximicen la recuperación funcional de las y los pacientes. En este contexto, el desarrollo de tecnologías asistivas, es decir, sistemas interactivos de software que permiten apoyar a personas con alguna dificultad de accesibilidad en realizar satisfactoriamente distintas tareas, surge como un enfoque prometedor. Así, el seguir procesos de diseño participativo y la integración de saberes y prácticas de distintas áreas de conocimiento —en este caso, ingeniería de software y fisiatría— permite explorar y consolidar redes de colaboración virtuosas que se traducen en el desarrollo de nuevos prototipos que permitan mejorar la efectividad de tratamientos médicos, así como la calidad de vida de pacientes y sus redes sociales de apoyo y cuidado.

Lo que comenzó hacia fines del año 2022 como un café informal entre dos científicos trabajando en áreas de especialización muy distantes entre sí —y que muy recientemente se consolidó en la firma de un acuerdo formal de colaboración entre la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile (FCFM) y COANIQUEM— se ve reflejado en el desarrollo de una familia de prototipos al servicio de las y los niños portadores de quemaduras, integrando colaborativamente la experiencia técnica de múltiples equipos en la intersección de ingeniería y medicina [1]. En este artículo presentamos estas herramientas, describiendo los objetivos que persiguen y las principales decisiones de diseño que tomamos en cada caso.

Reduciendo experiencias de dolor agudo en el tratamiento de quemaduras

La cicatrización de heridas producidas por quemaduras requiere de un esfuerzo coordinado de equipos multidisciplinarios, pasando por rehabilitadores físicos, trabajadores sociales y sicólogos. Así, dada la complejidad de este proceso, se requieren múltiples procedimientos que están en la raíz de experiencias altamente dolorosas, como lo son las cirugías y la curación de heridas. Naturalmente, estos procedimientos —aun cuando se hagan con cuidado por parte de profesionales— son altamente propensos a una carga emocional significativamente negativa. Los enfoques tradicionales para lidiar con el manejo del dolor se sustentan en la administración de sedantes y fármacos, los que cuentan con numerosas restricciones de uso en menores de 18 años.

El dolor es una experiencia sensorial. En contextos de rehabilitación, los videojuegos han ido apareciendo sostenidamente como estrategias plausibles y potencialmente efectivas, en particular, dada su capacidad de deslocalizar la atención del paciente hacia experiencias más agradables que lleven a un estado de *flow* [2], esto es, un estado cognitivo caracterizado por la confluencia de una alta inmersión y foco en la ejecución de una determinada tarea. Asimismo, el estado actual de desarrollo de mercado de dispositivos para interactuar con aplicaciones en realidad virtual ha permitido su masificación, debido a costos cada vez más reducidos y mayores posibilidades para que desarrolladores de software puedan producir nuevas aplicaciones siguiendo este modo de interacción. Así, no resulta extraño que surjan nuevas líneas de producción en el dominio de experiencias terapéuticas interactivas [3]. Por ejemplo, en el caso particular de la curación avanzada de heridas secundarias a quemaduras, existe el potencial de explorar el uso de experiencias altamente interactivas mediadas en realidad virtual debido al potencial de inmersión y *flow* que pueden experimentar los pacientes, llevando así a un estado de desensibilización en el que pueden reenfocar su atención en un ambiente multimodal, desde un estímulo doloroso hacia una experiencia más agradable [4].

En esta primera familia de prototipos, desarrollamos videojuegos controlados en ambientes mediados por realidad virtual, con un fuerte componente de interactividad, en los cuales las y los pacientes puedan experimentar experiencias inmersivas que lleven a un estado de desensibilización (ver Figura 1) durante las sesiones de curación avanzada de heridas.

Para cumplir satisfactoriamente con los objetivos de este desarrollo, tuvimos que enfrentar dos desafíos técnicos principales. En primer lugar, es necesario mantener un ambiente

Lo que comenzó como un café informal entre dos científicos [...] se consolidó en un acuerdo formal FCFM–COANIQUEM y en una familia de prototipos al servicio de las y los niños con quemaduras.

estéril en la sala de curaciones (y los controles físicos son elementos extraños que se deben mantener lejos del área a curar). Asimismo, dado que la gran mayoría de estas heridas ocurren en el tren superior (brazos y manos), debimos explorar mecanismos no convencionales para asegurar una interacción fluida, lo que incluye repensar metáforas de navegación y control, que no dependan de los mandos clásicos, para operar la aplicación. Como respuesta, decidimos integrar mecanismos basados en el seguimiento de mirada, lo que se puede capturar con cámaras que actúan como un complemento de hardware que puede ser integrado en algunos modelos de cascos de realidad virtual actualmente disponibles en el mercado de masas. Esta decisión implica tener que abordar activamente el llamado "toque de Midas" [5], una compensación explícita que se debe introducir en los mecanismos de control en medios de interacción natural (por ejemplo, aquellos basados en tacto y gestos) para evitar el riesgo de detectar acciones involuntarias del usuario, como por ejemplo pestañear o mover rápidamente el ojo en forma de paneo por distintos puntos de la interfaz. En este trabajo nos inspiramos en el método propuesto por Velichkovsky et al. [6], el cual consiste en diferenciar distintos eventos de interacción producidos por una fijación ambiental (por ejemplo, el paneo panorámico sobre la pantalla, caracterizado por movimientos en visión periférica) y por la fijación local (por ejemplo, la atención fija sobre un punto específico de la interfaz de usuario, en el cual no se detectan movimientos oculares frecuentes).

Por otro lado, la rehabilitación física requiere de un alto nivel de compromiso por parte de los pacientes. Lo anterior se debe a que las distintas sesiones que conforman el plan terapéutico de trabajo son usualmente largas, tediosas, repetitivas y dolorosas. Esto último se ve aún más acentuado en pacientes jóvenes, dado que tienden a perder el foco más fácilmente. En efecto, de acuerdo con Lohse et al. [7], una falta de adherencia a los procesos terapéuticos frecuentemente resulta en una barrera significativa para la rehabilitación. Así pues, basándonos en avances recientes en las áreas de modelamiento de usuarios y sicología de jugadores, exploramos la generación de modelos autoadaptativos y personalizados

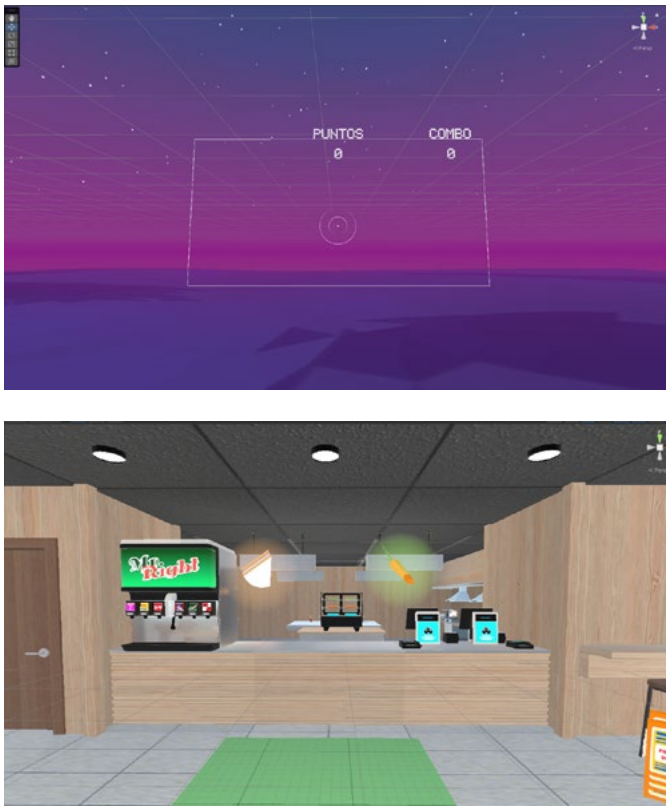


Figura 1 / Videojuegos mediados por realidad virtual para disminuir la intensidad de dolor en curación de heridas secundarias a quemaduras en tren superior de pacientes pediátricos.

para la selección de juegos y desafíos de interacción, tomando como base las distintas pistas contextuales, preferencias y perfiles de jugadores siguiendo la taxonomía BrainHex [8]. De esta manera, nuestra hipótesis es que las y los pacientes puedan sentirse comprometidos por un tiempo mayor, demostrando así mejor adherencia al tratamiento.

Videojuegos para el entrenamiento del rango articular

La rehabilitación mediada y moderada por videojuegos permite que los pacientes puedan mejorar su recuperación, tanto a nivel físico como mental, a través de la interacción sostenida con estímulos lúdicos que lleven a un estado de *flow*. Más precisamente, en la literatura médica los videojuegos, sobre todo aquellos clasificados como "videojuegos serios", han sido objeto de estudio para determinar su nivel de efectividad en lograr compromiso y adherencia a los procesos terapéuticos, en los que el paciente traslada su foco desde el dolor hacia una distracción inmersiva y presente, en la que potencialmente se logra un estado de desensibilización.

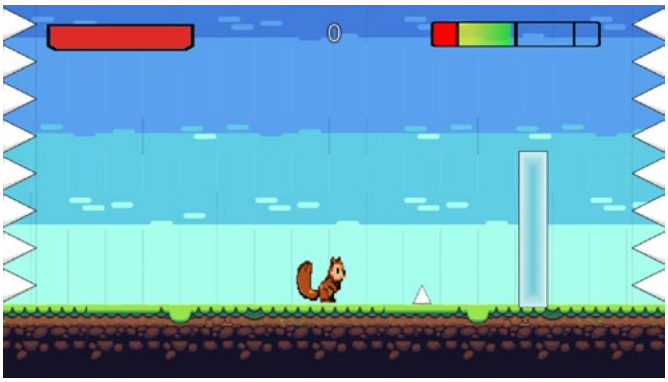


Figura 2 / *Exergames* para entrenar el rango articular como complemento a terapias físicas convencionales de pacientes portadores y secuelados de quemaduras.

En el estado actual de la práctica, los mecanismos de control en videojuegos se han diversificado, desde controles tradicionales como mouse, teclado y joysticks, hacia otros mecanismos de interacción natural como controles de movimiento, gestos y realidad virtual. En particular, es esta línea la que nos inspiró a explorar cómo podemos integrar el potencial de rehabilitación que ofrecen las interacciones altamente interactivas para lograr estados de *flow* y, simultáneamente, ejercitar activamente distintas partes del cuerpo, que son el centro de los procesos de rehabilitación física. En la literatura, a este tipo de videojuegos se les conoce como *exergames* (portmanteau de *exercise* y *video game*), en los que el objetivo principal consiste en integrar explícitamente la actividad física como parte de la experiencia de juego. Algunos ejemplos —ya clásicos— de exergames son las familias WiiFit y WiiSports de Nintendo, y Kinect Sports de Microsoft.

El estado de avance en el desarrollo de controles de videojuegos ha ido evolucionando para permitir la ejecución de interacciones naturales altamente fluidas y facilitar la detección de movimientos finos a través de sensores embebidos como giroscopios y acelerómetros. De esta manera, no es

Desarrollamos historias visuales interactivas en realidad virtual que [sitúan] al paciente en contextos cotidianos, tales como presentarse a sus compañeros en una sala de clases o participar de una entrevista de trabajo.

extraño pensar que podríamos capturar de manera pasiva distintas señales directamente del paciente, tales como funciones biométricas (como ritmo cardiaco y variación de pulso), así como otras propias de la biomecánica de distintas articulaciones como producto de la interacción física con el juego (como el ángulo de flexión del codo o la capacidad de rotación de la muñeca). En terapia física, se conoce como "rango articular" a la medida límite en la cual alguna parte del cuerpo puede moverse en torno a alguna articulación (como el codo, hombro, muñeca y rodilla), y es una de las métricas clave que se entrenan y monitorean en rehabilitación física por parte de los fisiatras.

En esta segunda familia de prototipos, desarrollamos una serie de videojuegos cortos de tipo exergame, capaces de entrenar de manera lúdica el rango articular de distintas partes del cuerpo críticas en la rehabilitación física de pacientes con secuelas de quemaduras: hombro, codo, muñeca y rodilla (ver Figura 2). De esta manera, buscamos generar un complemento a las terapias físicas convencionales que se basan en la ejercitación activa por parte de las y los pacientes mediante movilizaciones activas controladas y localizadas, pero que son propensas a una alta repetición y baja adherencia dada su monotonía. Así, nuestra hipótesis de base propone converger el alto poder inmersivo que puede tener una experiencia lúdica correctamente diseñada para el propósito y la población objetivo, junto a las restricciones técnicas y específicas del dominio que logren generar un entrenamiento de rango articular efectivo que perdure en el tiempo.

Uno de los desafíos principales para lograr nuestro propósito fue contar con una medida lo suficientemente precisa del movimiento, de tal forma que el equipo tratante pudiese monitorear la evolución del rango articular a lo largo del proceso terapéutico. Esto lo conseguimos al procesar las distintas señales capturadas por los controles de movimiento (en este caso, usamos los Joy-Cons de Nintendo Switch, que destacan por su tamaño y estética altamente valoradas por niñas y niños) y no comprometen significativamente la precisión en la captura e interpretación de señales. Asimismo,

siguiendo procesos de calibración y procesamiento de las distintas medidas de rotación y desplazamiento espacial, conseguimos desarrollar experiencias de juego altamente interactivas que permiten exportar medidas que pueden ser analizadas de manera clínica, posteriormente, por el equipo de fisiatras a cargo de la rehabilitación integral del paciente.

Apoyando la reinserción social de pacientes secuelados de quemaduras

El tratamiento de quemaduras es una tarea compleja y multidisciplinaria que va más allá de la recuperación física y la curación de heridas. En las etapas finales de este proceso, las y los pacientes (sobre todo aquellos pediátricos) deben recomponer sus herramientas sociales y emocionales para poder hacer frente a condiciones potencialmente adversas que puedan encontrar al reinsertarse plenamente en la sociedad [9]. Así pues, construyendo sobre casos exitosos en el uso de estrategias mediadas y moderadas por realidad virtual en contextos de simulación activa (como, por ejemplo, el tratamiento de fobias), en esta línea de trabajo buscamos desarrollar historias interactivas y lúdicas que permitan apoyar activamente a las y los pacientes en su reinserción social una vez hayan completado sus procesos de rehabilitación física. Estas aplicaciones, que actúan como un complemento a la terapia cognitivo-conductual (a cargo de los sicólogos clínicos), pretenden ser un punto de encuentro basado en el juego, la inmersión y los estados de *flow* que son propios de los videojuegos en realidad virtual.

En esta familia de prototipos, desarrollamos una serie de historias visuales interactivas en realidad virtual que permitan al paciente situarse —inmersivamente— en distintos contextos cotidianos de interacción social, tales como presentarse a sus compañeros en una sala de clases o participar de una entrevista de trabajo (ver Figura 3). Estos escenarios y dinámicas de interacción interpersonal fueron concebidos en conjunto con el equipo multidisciplinario de profesionales a cargo de los procesos de rehabilitación, en particular fisiatras y sicólogos especialistas en terapia de exposición controlada, construyendo así experiencias significativas, con un alto potencial de utilidad y adopción, y que permitan ser un aporte en el proceso de rehabilitación.

Un desafío técnico importante es integrar de manera efectiva una serie de mecanismos que articulen armónicamente las experiencias de involucramiento, presencia e inmersión a lo largo de la simulación interactiva. Para ello, nos basamos explícitamente en constructos derivados de la teoría de la autodeterminación [10] para diseñar metáforas de interacción potencialmente efectivas que motiven y aumenten el compromiso de los usuarios. En particular, nos basamos en

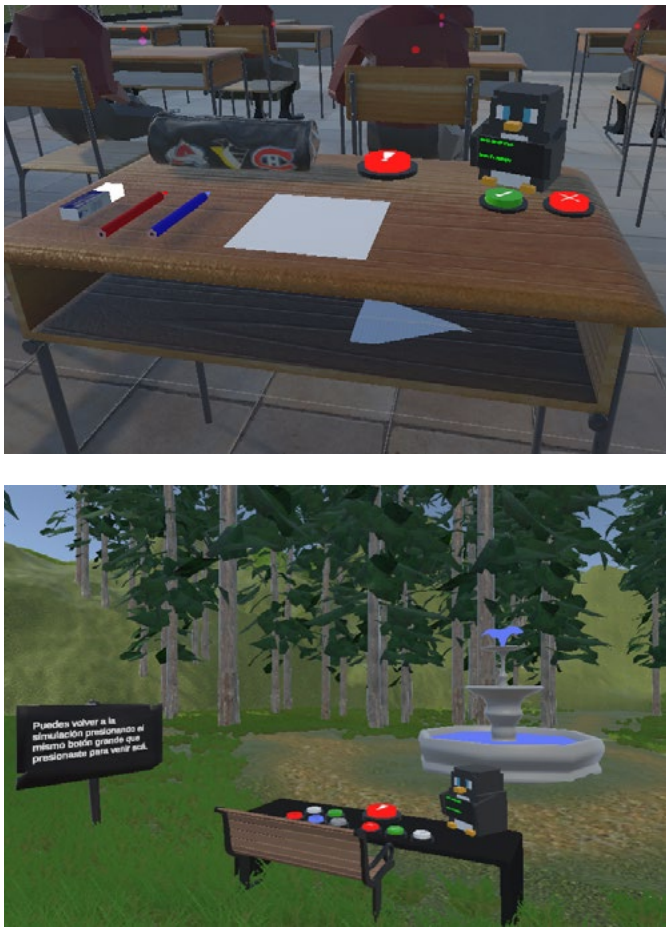


Figura 3 / Simulaciones de escenarios de interacción social en realidad virtual para promover la reinserción de pacientes una vez que completen su rehabilitación física.

el trabajo de Grasse et al. [11], quienes integraron los constructos principales de esta teoría (es decir, autonomía, competencia y conexión) para proponer mecanismos específicos en realidad virtual que medien y moderen la interacción en contextos de novelas narrativas. Asimismo, otra componente crítica para asegurar una interacción fluida y con sentido para el usuario es el desarrollo de zonas seguras, relajación y confort para controlar estados en los que el/la paciente pueda sentir indicios de estrés o ansiedad como producto de la exposición a estímulos aversivos o situaciones sociales que generen incomodidad. Esto lo conseguimos inspirándonos en la técnica de *shirin-yoku* (o "baño de bosque", una práctica japonesa de relajación que consiste en la inmersión plena buscando la conexión sensorial con la naturaleza), la cual ha presentado resultados potencialmente efectivos en su potencial de ser mediada por realidad virtual inmersiva [12], integrando así activamente la presencia del sicólogo clínico en la zona segura como un apoyo eventual.

Conclusión

En este artículo se presentaron tres familias de prototipos orientadas a apoyar el proceso de rehabilitación de pacientes con quemaduras en diferentes etapas de su evolución, abarcando desde la fase aguda hasta la fase secuelar. Estas propuestas comprenden diversas instancias del tratamiento, que incluyen la curación de heridas, la realización de ejercicios destinados a la mantención y recuperación del rango de movimiento en distintas articulaciones, así como la reinserción social como componente esencial del proceso rehabilitador.

Los desarrollos descritos fueron el resultado de una colaboración estrecha y sinérgica entre memoristas de Ingeniería Civil en Computación del DCC y el equipo profesional de rehabilitación de COANIQUEM, quienes participaron activamente en las distintas fases del proceso. Este enfoque evidencia la relevancia del trabajo multidisciplinario y del diseño participativo en la creación de soluciones tecnológicas potencialmente efectivas para dominios de aplicación específicos. En consecuencia, se promueve la integración del conocimiento técnico y clínico, generando nuevas líneas de investigación, desarrollo e innovación centradas en el usuario, con un impacto tangible en la sociedad. B

El tratamiento de quemaduras es una tarea compleja y multidisciplinaria que va más allá de la recuperación física y la curación de heridas.

Agradecimientos

Quisiéramos reconocer y agradecer el trabajo de Éric Contreras, Pablo Gutiérrez, Mario Recabarren, Franz Widerstrom, Ignacio Fuentes, Benjamín Bustos, Gonzalo Cartes y Luciano Providel, quienes con mucho entusiasmo y dedicación contribuyeron al desarrollo de los prototipos descritos en este artículo, como parte de sus proyectos para optar al título de Ingeniero Civil en Computación.

Referencias

[1] Corporación de Ayuda al Niño Quemado (COANIQUEM). 2022. Memoria Anual: Innovación y Rehabilitación Integral en Pacientes Pediátricos con Quemaduras. Santiago, Chile.

[2] Bruno Bonnechère, Bart Jansen, Lubos Omelina & Serge Van Sint Jan. 2016. The use of commercial video games in rehabilitation: A systematic review. *International Journal of Rehabilitation Research* 39(4):277–290.

[3] Dmitriy Viderman, Karina Tapinova, Mukhit Dossov, Serik Seitenov & Yerkin G. Abdildin. 2023. Virtual reality for pain management: An umbrella review. *Frontiers in Medicine* 10.

[4] Rania R. Ali, Ali Osman Selim, Mohamed A. Abdel Ghafar, Osama Ragaa Abdelraouf & Olfat Ibrahim Ali. 2022. Virtual reality as a pain distractor during physical rehabilitation in pediatric burns. *Burns* 48(2):303–308.

[5] Robert J. K. Jacob. 1990. What you look is what you get: Eye movement-based interaction techniques. En: *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'90)*. ACM Press, pp. 11–18.

[6] Boris Velichkovsky, Mikhail A. Rumyantsev & Mikhail A. Morozov. 2014. New solution to the Midas Touch problem: Identification of visual commands via extraction of focal fixations. Procedia Computer Science 39.

[7] Keith Lohse, Navid Shirzad, Alida Verster, Nicola Hodges & H. F. Machiel Van der Loos. 2017. Video games and rehabilitation: Using design principles to enhance engagement in physical therapy. *Journal of Neurologic Physical Therapy* 37(4):166–175.

[8] Lennart E. Nacke, Chris Bateman & Regan L. Mandryk. 2014. BrainHex: A neurobiological gamer typology survey. *Entertainment Computing* 5(1):55–62.

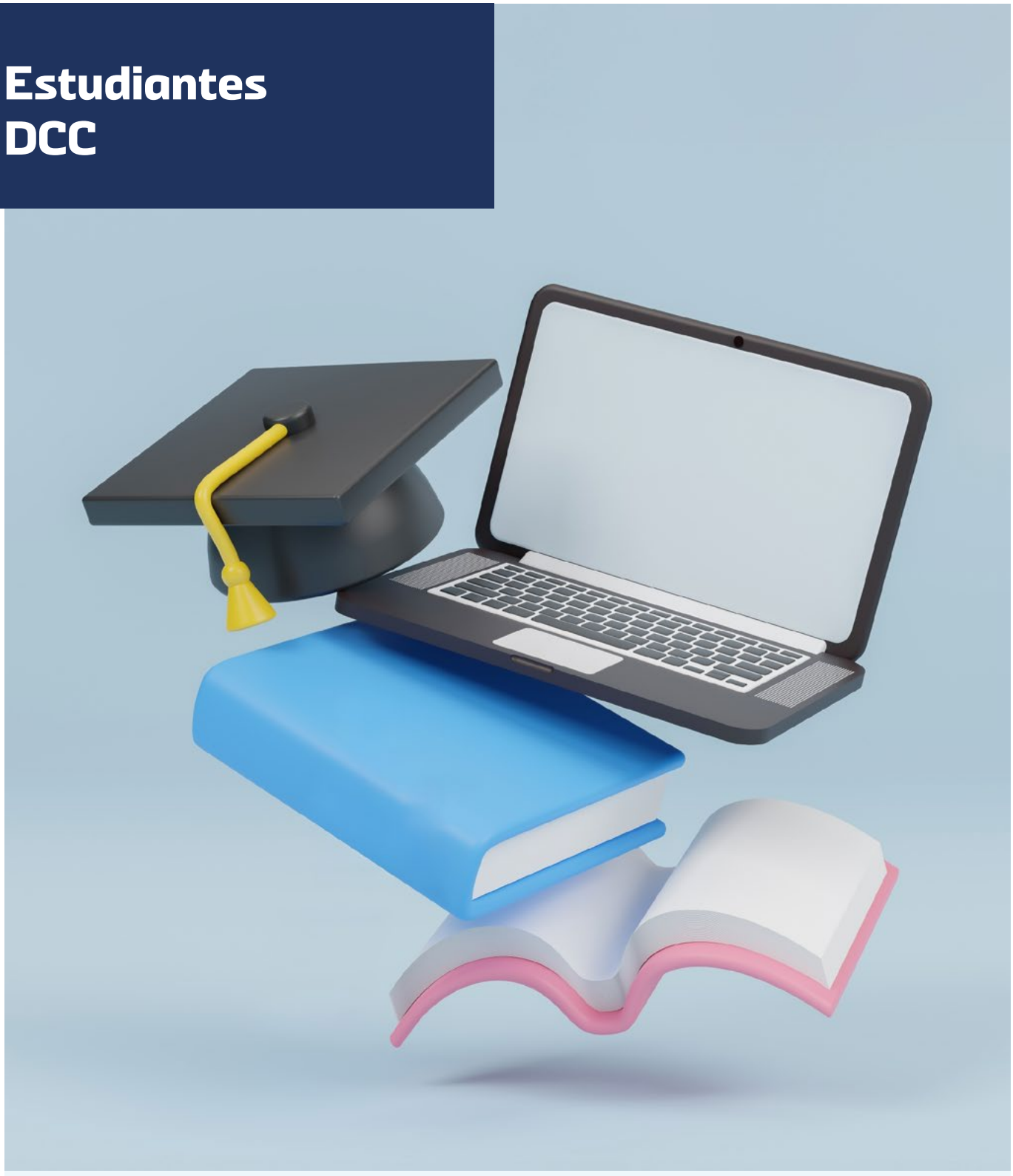
[9] Thomas D. Parsons & Albert A. Rizzo. 2008. Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: A meta-analysis. *Journal of Behavior Therapy and Experimental Psychiatry* 39(3):250–261.

[10] Richard M. Ryan & Edward L. Deci. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist* 55(1):68–78.

[11] Katelyn M. Grasse, Max Kreminski, Noah Wardrip-Fruin, Michael Mateas & Edward F. Melcer. 2022. Using self-determination theory to explore enjoyment of educational interactive narrative games: A case study of Academical. *Frontiers in Virtual Reality* 3.

[12] Rachel Masters, Jalynn Nicoly, Vidya Gaddy, Victoria Interrante & Francisco R. Ortega. 2024. The impact of natural realism on the restorative quality of virtual reality forest bathing. *ACM Transactions on Applied Perception* 22(1), artículo 3.

Estudiantes DCC



En esta sección de la Revista estudiantes recientemente *graduadxs* 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

Topological data analysis for classification of noisy and high-dimensional datasets

Estudiante
Rolando Kindelan

Profesores guías
Nancy Hitschfeld Kahler y
Mauricio Cerda

Nací en Cuba, donde estudié Ingeniería en Ciencias Informáticas y luego realicé un magíster en la misma Universidad de las Ciencias Informáticas. Durante más de siete años trabajé en la industria, en proyectos de desarrollo de software y equipos médicos para el sistema de salud cubano. Esa experiencia me enseñó lo valioso que es cuando la investigación se traduce en soluciones concretas para mejorar la vida de las personas.

En 2018 llegué a Chile para comenzar el Doctorado en Computación en la Universidad de Chile, bajo la guía de Nancy Hitschfeld Kahler y Mauricio Cerda. Fue un cambio enorme: adaptarme a un nuevo país, integrarme a una comunidad académica distinta y, al mismo tiempo, retomar la vida de estudiante después de tantos años en el mundo laboral.

Mi tesis, titulada *"Topological Data Analysis for Classification of Noisy and High-Dimensional Datasets"*, se centró en el Análisis Topológico de Datos (TDA), un enfoque emergente que aplica herramientas de la topología y la geometría para estudiar la "forma" de los datos. Me gusta explicarlo con una imagen sencilla: pensemos en un cardumen de peces o en una bandada de estorninos. Si estamos dentro, sólo vemos a los individuos cercanos; lo global se pierde. El aprendizaje automático tradicional funciona un poco así: capta patrones locales. El TDA, en cambio, nos permite mirar desde dentro y desde fuera a la vez, descubriendo tanto estructuras locales como patrones globales que guían el movimiento del conjunto.

Gracias a esta capacidad, el TDA suele usarse como complemento en los procesos de análisis de datos. Sin embargo, en mi investigación quisimos ir más allá y demostrar que puede ser una herramienta central para enfrentar desafíos como datos con ruido, valores faltantes, clases desbalanceadas o etiquetas incorrectas. En muchos de estos problemas, lo que realmente importa es entender bien la estructura de los datos, y ahí el TDA ofrece una perspectiva única.

Un hallazgo clave fue notar que muchas soluciones tradicionales dependen de *grafos de vecinos cercanos*, que representan sólo relaciones de a pares y no siempre reflejan la forma real de los datos. Por eso propusimos reemplazarlos por estructuras más expresivas, capaces de capturar relaciones de orden superior, como triángulos o cavidades. Esto abrió nuevas posibilidades para diseñar métodos de clasificación más robustos y aplicables a una amplia gama de escenarios.

Otro gran desafío fue el costo computacional: los objetos que maneja el TDA crecen rápidamente y son difíciles de almacenar y procesar. Para resolverlo, diseñamos estructuras de datos compactas que reducen drásticamente el uso de memoria sin perder eficiencia. También introdujimos nuevas formas de comparar de manera rápida y confiable las "huellas topológicas" de los datos, lo que permitió escalar los métodos a conjuntos mucho más grandes.

Más allá de los resultados técnicos, este doctorado fue para mí un viaje de transformación personal. Aprendí a explicar ideas complejas en un lenguaje sencillo, a trabajar en colaboración con investigadores de distintos países y a no rendirme frente a problemas que parecían imposibles. Hoy miro hacia atrás con gratitud: por mi familia, que me acompañó en cada paso; por mis profesores, que me guiaron con paciencia y rigor; y por la comunidad del DCC, que me recibió con generosidad desde el primer día.

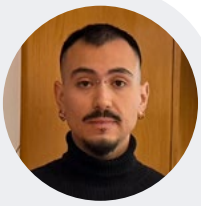
Actualmente me desempeño como Staff R&D Engineer en Synopsys Chile Innovation Center, en el grupo de Soluciones de Procesamiento Distribuido, ayudando a que los productos de la compañía escalen en eficiencia y rendimiento. Además, codicto un curso de Introducción al Análisis Topológico de Datos en la Universidad Católica, y mi meta es poder impartirlo este año acá en nuestro DCC. Con ello espero seguir difundiendo esta área emergente y contribuir a formar nuevas generaciones de investigadores que exploren cómo la topología puede ayudarnos a entender mejor los datos que gobiernan nuestro mundo.

Tesis de doctorado

A study on repetitiveness measures for strings

Estudiante
Cristian Urbina

Profesor guía
Gonzalo Navarro



Estudié Licenciatura en Ciencia de la Computación en la Universidad de Santiago de Chile, donde rápidamente desarrollé un gusto por los aspectos teóricos de la Ciencia de la Computación, en especial Teoría de Conjuntos, Matemática Discreta y Teoría de la Computación.

Poco después de terminar mi pregrado, comencé mi doctorado en la Universidad de Chile financiado por una beca ANID Doctorado Nacional, donde decidí realizar mi tesis con el profesor Gonzalo Navarro.

La problemática en que nos concentramos para mi doctorado fue la siguiente: Hoy en día, existen enormes colecciones de texto donde gran parte de los datos son repetidos, o muy similares entre sí. Esto es especialmente evidente en colecciones de genomas en bioinformática, donde dos genomas humanos coinciden en más del 99% de su contenido. Estas colecciones son tan grandes que deben ser comprimidas para poder ser manejables. Idealmente se busca mantener la posibilidad de responder consultas del texto original en espacio comprimido.

Los compresores e índices de texto tradicionales basados únicamente en la entropía de Shannon son capaces de explotar las frecuencias relativas de los símbolos como factor de compresión, pero no pueden capturar la repetitividad. Por ejemplo, la entropía de Shannon nos dice que para representar la concatenación de n copias del texto 01, necesitamos $2n$ bits. Explotando la repetitividad de este texto, nos damos cuenta de que basta con almacenar 01 con 2 bits, y el entero n usando $log(n)$ bits. Esto hace evidente que la entropía de Shannon no es una buena medida de compresibilidad para colecciones altamente repetitivas.

Mi tesis se titula "A Study on Repetitiveness Measures for Strings" y estudia, desde un punto de vista principalmente teórico, las distintas medidas de repetitividad que se han propuesto como alternativa a la entropía de Shannon. Muchas de estas medidas están asociadas al tamaño de compresores ampliamente utilizados en la práctica, como variantes del algoritmo de compresión Lempel-Ziv, el *run-length encoding* de la transformada de Burrows-Wheeler, u otras ideas basadas en compresión usando gramáticas libres de contexto.

En concreto, nuestras contribuciones son las siguientes:

1. Estudiamos propiedades combinatoriales de varias medidas de repetitividad del estado del arte, especialmente la transformada de Burrows-Wheeler. En específico, estudiamos qué tan robusta es esta medida de repetitividad cuando los textos son dinámicos, es decir, pueden cambiar en el tiempo.

2. Introdujimos nuevas medidas de repetitividad basadas en la noción de morfismo sobre textos. Mostramos que estas medidas son competitivas, e incluso ofrecen la posibilidad de romper lo que muchos consideran cotas inferiores para la repetitividad.

3. Extendimos las gramáticas libres de contexto utilizadas para compresión con nuevos tipos de reglas, obteniendo de esta forma una nueva representación comprimida más poderosa en términos de espacio, y que retiene gran parte de la funcionalidad que ofrecen las gramáticas para indexar textos comprimidos.

4. Generalizamos medidas de repetitividad de textos en una dimensión, a textos en d-dimensiones. Esto es importante porque existen colecciones de datos multidimensionales (como matrices, grafos, coordenadas, entre otros) que son altamente repetitivas, pero el estudio de la repetitividad hasta hace poco se concentraba únicamente en textos de una dimensión.

Mi experiencia en el doctorado fue muy positiva. Tuve la posibilidad de colaborar y conocer personas brillantes de distintas partes del mundo, lo cual me mantiene motivado a seguir mejorando como investigador. También pude realizar dos estadías de investigación en la Universidad de Palermo en Italia, y presentar varios artículos en conferencias internacionales. Recientemente participé en la publicación de dos artículos en las conferencias MFCS 2025 y SPIRE 2025. En esta última, obtuvimos el *Best Paper Award*.

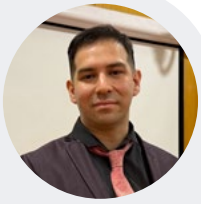
Actualmente estoy postulando y esperando resultados de becas para realizar un postdoctorado en el extranjero. A corto plazo pienso continuar investigando medidas de repetitividad, y de a poco empezar a explorar otras líneas de investigación que me puedan ayudar a desarrollar nuevas ideas.

Tesis de doctorado

Polygonal/polyhedral mesh generation from Delaunay tessellations

Estudiante
Sergio Salinas Fernández

Profesora guía
Nancy Hitschfeld Kahler



Entré al doctorado por una noticia que me llegó al celular, mientras trabajaba de programador *full stack*. Era una noticia de las investigaciones de la doctora Nancy Hitschfeld, profesora del DCC. Las encontré fascinantes y me contacté por correo para ver si podía cooperar con ella. Amablemente me coordinó una reunión presencial, donde me explicó sus investigaciones y me recomendó entrar al doctorado a trabajar con ella. Fue así como nació mi investigación doctoral, el algoritmo Polylla para la generación de mallas poligonales.

Cuando buscaba temas, quería trabajar en algo altamente matemático, pero también práctico; ahí descubrí la generación de mallas poligonales. Para explicarlo fácil: Imagine que se quiere simular si un edificio soporta un terremoto. Este tipo de problemas requiere que se pueda representar una geometría, en este caso un edificio, en una computadora para aplicar algún método numérico sobre el objeto. Para ello se usan varios elementos planos iguales, como triángulos o cuadrados conectados, sin que se superpongan, que representan la geometría—básicamente "un mosaico"—pero siempre repitiendo la misma figura, y sobre esta geometría se le aplica un método numérico que haga la simulación. Pero existe un *trade-off*: si se usan muchos elementos, la simulación va a ser más lenta en tiempo y va a ocupar más memoria, pero va a ser más precisa; si se ocupan pocos elementos va a ser más rápida pero no tan útil.

Justo cuando entré, un método numérico relativamente nuevo estaba destacando: el Virtual Element Method (VEM). Este permite usar cualquier figura como elemento base, no sólo triángulos, y además no tienen que ser todas iguales. Pero no existía ningún algoritmo que hiciera una malla poligonal para el VEM, así que con Nancy nos propusimos resolver el problema creándolo nosotros mismos. De ahí surgió Polylla (*POLYgonaL meshing aLgorithm bAsed on terminal-edge regions*). Con Polylla, no sólo desarrollamos un nuevo tipo de generador de mallas poligonales, también ampliamos la investigación proponiendo métodos para generarlas usando aceleración con GPU, para más rapidez, y estructuras compactas, para usar menos memoria.

Sobre el doctorado, fue una experiencia bastante estresante en un inicio. Entré al doctorado sin magíster por lo que no tenía experiencia en investigación. Mi pregrado fue Licenciatura en Ciencias de la Computación en la Universidad de Santiago, cuando me contacté con Nancy, coincidió también que me había ganado varios premios por mi desempeño académico, por lo que lo tomaron en cuenta cuando aceptaron mi postulación.

Al entrar no sabía escribir publicaciones, ni demostrar matemáticamente algoritmos y nunca había trabajado con geometría computacional, ni temas relacionados. Muchas veces intenté renunciar y abandonar todo. Pero la profesora Nancy siempre fue tan amable y cooperativa conmigo; ella me ayudó a mitigar cualquier problema que tuve y me ayudó mucho a crecer como persona y profesionalmente.

Una vez que realizamos nuestra primera publicación todo se aligeró bastante, ya que la investigación captó la atención de investigadores internacionales, que nos dieron un excelente *feedback* sobre cómo seguir investigando, como también de investigadores nacionales que nos ayudaron a sacar más publicaciones. Incluso dentro del departamento la investigación hizo ruido y nos empezaron a llegar alumnos que querían trabajar con nosotros. Ya al final del doctorado me dedicaba más a dirigir investigaciones con los alumnos, que tener que escribirlas. Además, gracias al doctorado, pude hacer pasantías y viajar a conferencias internacionales, donde conocí sobre otras culturas, lo que logró que cambiara toda mi perspectiva de cómo veía el mundo.

Por desgracia, estar demasiado enfocado en el doctorado, me llevó a problemas familiares y de salud, por lo que decidí tomarme un descanso de la academia. Ahora trabajo como ingeniero de datos para la empresa francesa Alstom. Aún estoy decidiendo qué hacer con mi vida, pero agradezco mucho al doctorado y a Nancy por todas las oportunidades que me dieron. Especialmente a Nancy.

Tesis de doctorado

Multilingual hate speech detection

Estudiante
Aymé Arango

Profesores guías
Bárbara Poblete y Jorge Pérez

Tras finalizar mis estudios de pregrado en Ciencias de la Computación en la Universidad de Oriente, Cuba, comencé a explorar nuevas oportunidades de formación académica. Fue entonces, a través de compañeros de estudios, que supe de la posibilidad de realizar un doctorado en la Universidad de Chile, una alternativa que en mi entorno en Cuba no era muy conocida. Entre las distintas líneas de investigación que se desarrollan en el Departamento de Ciencias de la Computación (DCC), decidí orientarme hacia el área de Ciencia de Datos donde la profesora Bárbara Poblete me aceptó como su estudiante y, bajo su guía, desarrollé mi tesis doctoral sobre detección de discurso de odio en redes sociales.

La detección de discurso de odio en texto es un problema complejo de Procesamiento de Lenguaje Natural, en el que los modelos de aprendizaje automático deben ser capaces de sortear la ironía y las estrategias de oclusión léxica utilizadas para ocultar términos ofensivos. La mayoría de las soluciones y recursos disponibles en ese momento habían sido desarrollados para el idioma inglés, mientras que otros idiomas, como el español, estaban poco explorados.

En primer lugar, realizamos un análisis crítico del estado del arte en inglés, comprobando que los mejores resultados reportados estaban sobreestimados. Mostramos su limitada capacidad de generalización y las causas de este problema, como el uso de datos sesgados, lo que los hacía poco útiles en escenarios reales. Consideramos que el escenario *cross-lingual* puede entenderse como un caso extremo de generalización de modelos.

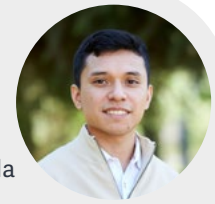
Diseñamos distintos modelos de aprendizaje automático capaces de clasificar texto en un idioma diferente al utilizado en el entrenamiento. Para ello, propusimos conjuntos de características multilingües específicamente diseñados para la tarea. El primero consistió en un conjunto de características extraídas de la metainformación de las redes sociales, como la popularidad de la publicación, el número de veces compartida o su alcance. Asimismo, propusimos un conjunto de *word embeddings* específicos para el dominio, y mostramos que estas representaciones, aunque simples, pueden capturar información más significativa en la detección de discurso de odio que modelos más complejos. El profesor Jorge Pérez aportó su experiencia en *deep learning* para aplicar técnicas avanzadas de modelamiento del lenguaje.

Uno de los principales desafíos para validar estos modelos fue la escasez de datos etiquetados. Por ejemplo, para el español sólo existían un par de conjuntos de datos centrados en la variante de España, sin considerar otras variedades. Ante esto, demostramos la importancia de la representación multicultural y construimos el primer conjunto de datos de discurso de odio originado en Chile, con la colaboración de un equipo multidisciplinario.

Aunque hubo momentos difíciles, el doctorado en la Universidad de Chile fue una experiencia enriquecedora, llena de aprendizaje y crecimiento personal. Tuve la oportunidad de conocer a muchas personas y una nueva cultura. Siempre conté con la ayuda de los profesores del DCC y de los investigadores del Instituto Milenio Fundamentos de los Datos (IMFD), del cual formé parte durante varios años.

Tesis de magíster

Reconocimiento de patrones repetitivos en imágenes de motivos de herencia cultural



Estudiante Sebastián Sepúlveda

Profesores guías Benjamín Bustos e Iván Sipirán

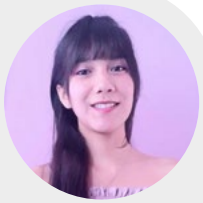
En este trabajo estudiamos y evaluamos herramientas computacionales para identificar automáticamente la posición en donde aparecen patrones de motivos de herencia cultural en la superficie de cerámica antigua. Para esto, utilizamos un conjunto digitalizado de 82 piezas de cerámica antigua pertenecientes al Museo Josefina Ramos de Cox, ubicado en Lima, Perú, para las cuales se anotaron manualmente la aparición de cada patrón relevante en la superficie de la cerámica. La tarea consiste en, dada la imagen de la superficie de una cerámica, detectar la aparición de cada instancia de algún patrón relevante. Una característica que tienen estos patrones es que, usualmente, aparecen en forma repetitiva a lo largo del contorno de la cerámica.

Para resolver este problema, evaluamos distintos algoritmos y métodos de detección y segmentación de objetos como YOLOv8, Retina-Net, Mask-RCNN, Faster-RCNN, Template Matching y Segment Anything. La evaluación consistió en dos estrategias distintas: una estándar, en donde los patrones relevantes se encontraban en el conjunto de entrenamiento, y una *zero-shot*, en donde los patrones relevantes no estaban en el conjunto de entrenamiento. Los resultados de la evaluación experimental muestran que YOLOv8 obtiene la mejor eficacia en el caso de la estrategia estándar, mientras que Retina-Net obtuvo el mejor resultado en el caso de la estrategia zero-shot.

Los principales resultados de esta investigación fueron publicados en el ACM Journal on Computing and Cultural Heritage. Este trabajo fue parcialmente financiado por el Proyecto ANID - Fondecyt Regular – N° 1230448.

Tesis de magíster

Análisis de la comprensión de modelos de lenguaje generativo en el comportamiento político chileno



Estudiante Vanessa Gaete

Profesores guías Andrés Abeliuk y Naim Bro

Los modelos de inteligencia artificial como ChatGPT han sorprendido por su capacidad para predecir comportamientos políticos en Estados Unidos. A partir de simples datos demográficos (como edad, género o nivel educativo), estos sistemas pueden generar respuestas que imitan de manera realista las opiniones humanas, lo que abre la posibilidad de reducir drásticamente los costos de las encuestas y transformar la investigación en ciencias sociales. Sin embargo, casi todos estos avances se han desarrollado y probado en contextos estadounidenses, dejando abierta una pregunta clave: ¿funcionan igual en otras sociedades?

Esta tesis es el primer estudio en Chile que evalúa si los modelos de lenguaje grandes pueden reproducir patrones de opinión y comportamiento político en la población chilena, comparando su desempeño con el obtenido en Estados Unidos. El objetivo fue determinar si estas herramientas podrían servir como métodos de predicción confiables para eventos políticos locales y qué estrategias permiten mejorar su rendimiento.

Se pusieron a prueba cuatro modelos (ChatGPT-4, ChatGPT-3.5, Llama-2-13b y Mistral) en tres escenarios: la elección presidencial chilena de 2021, el plebiscito constitucional de 2022 y las actitudes frente al aborto. Los resultados se contrastaron con experimentos equivalentes realizados con datos estadounidenses. Para ello se utilizaron encuestas públicas: las del Centro de Estudios Públicos (CEP) en Chile y las del American National Election Studies (ANES) en Estados Unidos.

Los resultados fueron claros: ninguno de los modelos logró realizar predicciones precisas en el contexto chileno, mostrando un desempeño consistentemente inferior al observado en Estados Unidos. Además, se detectaron sesgos significativos: en Chile, los modelos tuvieron más dificultad para predecir las opiniones de las mujeres que de los hombres, algo que no ocurrió en el caso estadounidense. En ambos países, las respuestas fueron más acertadas entre personas no religiosas (ateas o agnósticas) que entre quienes se identifican con una religión.

En conjunto, los resultados evidencian que, aunque la inteligencia artificial promete transformar la investigación social, su aplicabilidad fuera del contexto estadounidense sigue siendo limitada. Por lo tanto, es necesario desarrollar modelos y metodologías más inclusivos y contextualizados a la diversidad cultural y social antes de confiarles la tarea de predecir el pulso político de nuestras sociedades.

Memoria de pregrado

Aplicación web para diseñar bases de datos

Estudiante Matías López

Profesores guías Aidan Hogan y Sebastián Ferrada

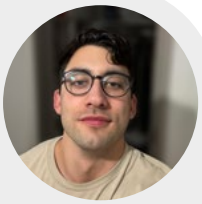
Un tema desafiante para los estudiantes de Bases de Datos (CC3201) es aprender cómo diseñar un modelo conceptual de un dominio usando diagramas Entidad–Relación (E–R). Tales diagramas capturan las entidades del dominio (p.ej., usuario, libros), sus atributos (p.ej., RUTs, nombres, títulos) y las relaciones entre entidades (p.ej., lee, califica, vende), para luego definir el esquema de la base de datos. La forma tradicional de hacer esta tarea era dibujar diagramas en papel, algo que no permitía recibir *feedback* hasta la evaluación de entrega final.

El objetivo del trabajo de título de Matías López —coguiado por Sebastián Ferrada y Aidan Hogan— fue diseñar, implementar y evaluar un sistema en línea que permite a los estudiantes de Bases de Datos definir, validar y visualizar diagramas E–R y así, recibir *feedback* inmediato sobre sus avances. Se diseñó un nuevo lenguaje y sintaxis para definir diagramas E–R, y se implementó una aplicación web que permite validar, parsear, detectar errores y visualizar estos diagramas.

Este trabajo fue publicado en el *International Workshop on Data Systems Education*, que forma parte de SIGMOD 2024: la conferencia más importante en bases de datos internacionalmente. Otra memorista, Kathleen Köhler, guiada por Matías Toro y Aidan Hogan, siguió el trabajo: ha implementado varias extensiones que incluyen la edición colaborativa en vivo de los diagramas E–R. La aplicación web (ERDoc) está disponible en línea (https://erdoc.dcc.uchile.cl/), ha sido usada por cientos de estudiantes del curso CC3201, y puede ser usada en otras universidades u otros contextos.

Memoria de pregrado

Implementación de algoritmos subóptimos para reordering en síntesis de circuitos integrados para Synopsys



Estudiante Diego Ruiz

Profesor guía Gonzalo Navarro

La industria de los circuitos integrados experimenta un rápido crecimiento y desempeña un papel fundamental en el avance tecnológico mundial. Synopsys, empresa líder en la industria, emplea la técnica de Scan Testing para verificar la funcionalidad de los circuitos integrados después de la fabricación, a través del proceso de Scan Insertion. Este proceso implica la incorporación de circuitos adicionales, incluida la conformación de una cadena entre componentes del circuito. En particular, el componente esencial de Scan Insertion, conocido como Reordering, se centra en optimizar la conectividad de esta cadena, buscando minimizar su largo. Este desafío se aborda como una instancia particular del Problema del Vendedor Viajero, conocido por ser NP-Completo.

Aunque los algoritmos actuales de Synopsys son funcionales, tienen un amplio margen de mejora y distan del estándar académico. Tras revisar heurísticas desarrolladas en la academia para el Problema del Vendedor Viajero, identificamos el algoritmo Lin-Kernighan con mejoras de Keld Helsgaun como prometedor para implementar en Synopsys. La implementación del algoritmo muestra resultados notables en la mejora de la optimalidad y tiempos de ejecución prometedores. Se planea que el nuevo algoritmo sea adoptado como el estándar en la herramienta de diseño de Synopsys.

Esta memoria resulta particularmente interesante por resolver un problema real, que se traduce en un problema algorítmico abstracto. Este problema, no trivial de resolver directamente, se aborda con heurísticas conocidas pero desafiantes para implementar. Finalmente, la solución desarrollada resulta exitosa en la práctica. Es un perfecto ejemplo de transferencia tecnológica de la académica a la industria.

Memoria de pregrado

Requerimientos geométricos de modelos hidrogeológicos de cuencas afectadas por megasequía: Caso de estudio cuenca del Limarí

Estudiante Antonio Torga

Profesores guías Nancy Hitschfeld Kahler, Pedro Sanzana y Felipe Troncoso

En el contexto del proyecto Fondecyt 1241596 dedicado al desarrollo de nuevos algoritmos para aplicaciones en ciencia e ingeniería computacional se genera una colaboración interdisciplinaria entre las ciencias de la computación e hidrología. Un primer resultado fue el software GeoLinkage, plugin del sistema de información geográfica GRASS GIS, encargado de generar automáticamente el archivo de "enlace" para la integración de un modelo hidrológico superficial, WEAP, con uno subterráneo, MODFLOW. La creación manual de estos archivos es un proceso engorroso, que toma a un modelador horas en un software GIS. Además, la creación "a mano" es propensa a errores, por lo que la asistencia computacional es crucial para generar resultados correctos que no provoquen una pérdida de flujo de agua entre los modelos.

GeoLinkage automatiza correctamente el proceso de creación de este archivo de enlace, reduciendo de horas a minutos, permitiendo además la iteración rápida de este archivo. Sin embargo, GeoLinkage no es capaz de diagnosticar errores en el archivo de enlace resultante, los cuales pueden ser generados por GeoLinkage, ya que hereda estos errores de sus archivos de entrada. Con el fin de llenar ese vacío, este proyecto de memoria actualiza el software GeoLinkage y lo extiende con el nuevo módulo GeoChecker, que ejecuta un chequeo automático del archivo resultante proveyendo visualizaciones de los errores encontrados y reportes detallando su magnitud y su causa.

El error geométrico que se buscaba diagnosticar en este proyecto es la superposición de elementos del modelo subterráneo (MODFLOW) en el archivo de enlace que no estuviera respaldada por una conexión en el modelo superficial (WEAP). Un ejemplo de este tipo de errores sería la superposición de una cuenca hidrológica o sectores de riego sobre un acuífero. Generalmente esto implicaría una comunicación de flujo entre ambos sistemas, sin embargo, si no existe una conexión entre ellos en WEAP el flujo se pierde, lo que induce errores en el balance hídrico completo. La actualización llevada a cabo, que incluye el nuevo módulo GeoChecker, diagnostica estos problemas, entregando al equipo de modelación la información necesaria para poder rectificar posibles problemas de enlace.

GeoChecker es un módulo añadido a GeoLinkage, y puede ser activado desde cualquiera de las interfaces de este programa (posee una interfaz en GRASS GIS y una de línea de comandos). Este trabajo fue presentado en la sesión de Hidroinformática de la asamblea general de la European Geosciences Union 2025, y también fue presentado en el Congreso de Hidráulica 2025 de la Sociedad Chilena de Ingeniería Hidráulica.

UNIVERSIDAD DE CHILE

dcc CIENCIAS DE LA COMPUTACIÓN UNIVERSIDAD DE CHILE

NUEVO

Diploma de Postítulo en Protección de Datos Personales

Desarrolla soluciones para gestionar datos personales de forma segura, ética y conforme a la nueva ley.

El diploma forma a profesionales en la implementación de soluciones que permitan dar cumplimiento efectivo a las exigencias de la nueva normativa chilena de protección de datos personales. Aborda la protección de datos desde una perspectiva aplicada, integrando técnicas de seguridad de la información, privacidad por diseño, anonimización y privacidad diferencial, gobernanza de datos y evaluación de equidad en sistemas de aprendizaje automático. Los contenidos legales se tratan como un marco de referencia operativo para la toma de decisiones de diseño e implementación de sistemas.

Inicio:
Agosto 2026

Modalidad:
Online sincrónico

COORDINACIÓN ACADÉMICA

Matías Toro
Doctor en Ciencias de la Computación. Profesor Asistente del DCC, U. de Chile.

Federico Olmedo
Doctor en Ciencias de la Computación. Profesor Asistente del DCC, U. de Chile.

Más información

CONTACTO

capacita@dcc.uchile.cl

+56 9 8434 8251

dcc.uchile.cl/ec

Departamento de
Ciencias de la Computación

EDUCACIÓN CONTINUA DCC

Potencia tu carrera en tecnología

- Clases online sincrónicas
- Enfoque práctico
- Certificación Universidad de Chile

Más información

DIPLOMAS

Actualiza y potencia tu perfil profesional en áreas clave de la computación.

- → Ciencia e Ingeniería de Datos
- → Gestión de Proyectos Informáticos
- → Ingeniería de Software
- → Inteligencia Artificial
- → Protección de Datos Personales
- → Python Aplicado a la Ciencia de Datos
- → Tecnologías de la Información

BOOTCAMPS

Inicia tu carrera en desarrollo de software; no requiere experiencia previa en programación.

- → Diseño UX/UI
- → Desarrollo Frontend
- → Desarrollo Backend
- → Desarrollo de Aplicaciones Móviles

MAGÍSTER EN TECNOLOGÍAS DE LA INFORMACIÓN

Especialízate en el desarrollo, adopción y gestión de tecnologías de la información.

PROGRAMAS CORPORATIVOS

Capacita a tus equipos con programas diseñados a la medida.

CONTACTO

capacita@dcc.uchile.cl
+56 9 8434 8251
dcc.uchile.cl/ec

/educacioncontinuadcc

UNIVERSIDAD DE CHILE

dcc CIENCIAS DE LA COMPUTACIÓN UNIVERSIDAD DE CHILE

B