



REVISTA DEL DEPARTAMENTO DE CIENCIAS DE LA
COMPUTACIÓN DE LA UNIVERSIDAD DE CHILE

Bits

DE CIENCIA

PRIMER SEMESTRE 2019

COMPUTACIÓN Y DEMOCRACIA

LA DEMOCRACIA
EN LA ENCRUCIJADA
DIGITAL | Martin Hilbert

PERSONALIZACIÓN, POSTVERDAD
Y DEMOCRACIA:

EFFECTOS DE LAS
TECNOLOGÍAS DIGITALES
DE LA INFORMACIÓN EN
LA CIUDADANÍA | Leandro de Brasi

NAVEGANDO SOBRE CORRIENTES DE
DATOS Y CÓMO NO AHOGARSE EN EL
INTENTO | Alejandro Grez | Cristián Riveros



COMITÉ EDITORIAL

Pablo Barceló
Benjamín Bustos
Claudio Gutiérrez
Alejandro Hevia
Jocelyn Simmonds
Sergio Ochoa

EDITOR GENERAL

Federico Olmedo

EDITORA PERIODÍSTICA

Ana Gabriela Martínez

PERIODISTA

Karin Riquelme

DISEÑO

Puracomunicación

FOTOGRAFÍAS E IMÁGENES

Comunicaciones DCC

Revista BITS de Ciencia
del Departamento de Ciencias de la
Computación de la Facultad de Ciencias
Físicas y Matemáticas de la Universidad
de Chile se encuentra bajo Licencia
Creative Commons Atribución-
NoComercial-Compartir-Igual 3.0 Chile.
Basada en una obra en
www.dcc.uchile.cl



Revista Bits de Ciencia Nº17

ISSN 0718-8005 (versión impresa)

www.dcc.uchile.cl/revista

ISSN 0717-8013 (versión en línea)



Departamento de Ciencias de la Computación

Avda. Beauchef 851, 3º piso,
edificio norte

Santiago, Chile

837-0459 Santiago

www.dcc.uchile.cl

Fono 56-2-29780652

Fax 56-2-26895531

revista@dcc.uchile.cl

CONTENIDOS

01 EDITORIAL

| Federico Olmedo

ACTUALIDAD



02 INSTITUTO MILENIO FUNDAMENTOS DE LOS DATOS: A LA VANGUARDIA EN LA INVESTIGACIÓN DE LOS PROBLEMAS ESENCIALES EN TORNO A LOS DATOS

| Noemí Miranda

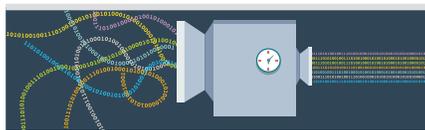
08 ¿DÓNDE ESTÁ LLOVIENDO AHORA? PREGÚNTALE A VISMET

| María Cecilia Bastarrica
| René Garreaud

14 EL TIEMPO DIRÁ...

| Éric Tanter

INVESTIGACIÓN DESTACADA



16 NAVEGANDO SOBRE CORRIENTES DE DATOS Y CÓMO NO AHOGARSE EN EL INTENTO

| Alejandro Grez
| Cristián Riveros

COMPUTACIÓN Y DEMOCRACIA



22 LA DEMOCRACIA EN LA ENCRUCIJADA DIGITAL

| Martín Hilbert

30 EFECTOS DE LAS TECNOLOGÍAS DIGITALES DE LA INFORMACIÓN EN LA CIUDADANÍA

| Leandro de Brasi

38 RANKINGS DE PERSONAS: EL CASO DEL SISTEMA DE CRÉDITO SOCIAL CHINO

| Claudio Gutiérrez

COMPUTACIÓN Y SOCIEDAD



42 ¿QUÉ ES GDPR?

| Luis Arancibia

48 LA HAZAÑA DEL RECUENTO PARALELO DEL PLEBISCITO DE 1988

| Juan Álvarez Rubio

CONVERSACIONES

54 ENTREVISTA A NICOLÁS LEHMANN

| Por Federico Olmedo

DOCTORADOS DEL DCC

56 | Francisco Gutiérrez | Daniel Moreno | Luis Silvestre



Luego de 10 años ininterrumpidos a cargo de nuestra Revista Bits de Ciencia, Pablo Barceló ha dejado su rol de editor para enfrentar nuevos desafíos profesionales en el Instituto Milenio Fundamentos de los Datos. En nombre de todo el equipo editorial queremos agradecerle su enorme esfuerzo y dedicación en este tiempo y desearle los mayores éxitos en su nuevo emprendimiento.

Personalmente, acepto con mucha felicidad y entusiasmo el desafío de mantener a nuestra Bits de Ciencia como un referente nacional en materia de divulgación científica y tecnológica.

Para este número de la Revista hemos elegido abordar como tema principal el rol de la computación y nuevas tecnologías en la democracia actual. En mayor o menor medida, todas las formas de democracia actuales fueron inspiradas por las primeras democracias originadas en la Antigua Grecia y Roma, en un contexto de tecnologías y formas de comunicación completamente distintas a las actuales. Sin tener que remontarse tantos siglos, pensemos que cuando Abraham Lincoln fue elegido Presidente de Estados Unidos (1860), los ciudadanos de California tardaron ¡cinco días! en enterarse del resultado de dicha elección, ya que los mensajes se llevaban a caballo.

En las últimas décadas las tecnologías de la información han evolucionado a una velocidad impensada, incluyendo los mecanismos de acceso y transmisión (piense, por ejemplo, en Internet) como también los mecanismos de procesamiento y sus respectivas aplicaciones (piense, por ejemplo, en el machine learning o aprendizaje de máquina). Este avance ha generado un impacto definitivamente imprevisto en los mecanismos de la democracia actual, impacto que trataremos de dilucidar en los distintos artículos de nuestra sección principal *Computación y Democracia*.

La sección comienza con un artículo de Martin Hilbert, donde se analiza el desajuste entre nuestra realidad tecnológica y nuestros procesos democráticos, poniendo especial énfasis en el uso del machine learning para abordar dicho desajuste. La sección continúa con un artículo de Leandro de Brasi, donde se discuten los efectos negativos de la personalización o “customización” de contenido que comúnmente llevan a cabo las tecnologías de la información actuales, y finalmente concluye con un artículo de Claudio Gutiérrez sobre el ranking de personas recientemente implementado por el gobierno chino.

A partir de este número, abriremos la Revista con una nueva sección *Actualidad*. En esta edición Noemí Miranda nos cuenta sobre la reciente fundación del Instituto Milenio Fundamentos de los Datos; los profesores de la Universidad de Chile, María Cecilia Bastarrica y René Garreaud se refieren al desarrollo de VisMet, un visualizador que centraliza la información meteorológica de las distintas estaciones a lo largo del país, y finalmente Éric Tanter, profesor del DCC de la misma universidad, se refiere a dos de sus trabajos que fueron recientemente premiados por su impacto.

La Revista continúa con sus secciones tradicionales. En la sección *Investigación Destacada*, Cristián Riveros y Alejandro Grez abordan el desafío que supone el procesamiento de streams continuos de datos. En la sección *Computación y Sociedad*, Luis Arancibia nos presenta el reciente Reglamento de Protección de Datos de la Unión Europea (GDPR) y Juan Álvarez nos cuenta sobre el sistema de recuento de votos paralelo que se desarrolló para el plebiscito de 1988. En la sección *Conversaciones* entrevistamos a Nicolás Lehmann, exalumno del DCC de la Universidad de Chile, actualmente trabajando en Silicon Valley. Finalmente, concluimos este número

de la Revista con una presentación de los últimos egresados de nuestro Programa de Doctorado.

Esperamos que disfruten este nuevo número de la Revista. Como siempre, no duden en enviar sus comentarios y consultas a revista@dcc.uchile.cl.

FEDERICO L MEDO

Editor General

Revista Bits de Ciencia





INSTITUTO MILENIO
FUNDAMENTOS DE
LOS DATOS:
A LA VANGUARDIA
EN LA INVESTIGACIÓN
DE LOS PROBLEMAS
ESENCIALES EN
TORNO A LOS
DATOS



**NOEMÍ
MIRANDA**

Periodista. Directora de Comunicaciones y Proyección del Medio del Instituto Milenio Fundamentos de los Datos. Especializada en comunicación de la ciencia y la tecnología, área en la que ha sido editora, profesora universitaria, organizadora y expositora en diversos congresos sobre la materia. noemi.miranda@imfd.cl



AL PRIMER WORKSHOP DEL INSTITUTO FUNDAMENTOS DE LOS DATOS, REALIZADO EN AGOSTO DE 2018 EN MACHALÍ, ASISTIERON ALREDEDOR DE 100 ACADÉMICOS Y ESTUDIANTES DE LAS DISTINTAS DISCIPLINAS Y UNIVERSIDADES QUE AGRUPA LA INSTITUCIÓN.

Financiado por la Iniciativa Científica Milenio, Fundamentos de los Datos es uno de los primeros centros multidisciplinarios de América Latina que realizará investigación de frontera en ciencia de datos con expertos provenientes de la ciencia política, las comunicaciones, las ciencias de la computación y la estadística. La meta es abordar los desafíos urgentes en materia de datos, desde su comprensión como compleja unidad informática, hasta sus usos e impacto en la sociedad.

En un mundo globalizado e hiperconectado, la información cumple un rol vital en ámbitos que van desde la administración de sistemas cruciales para el funcionamiento de un país a las interacciones que realizamos cotidianamente en redes sociales.

Se calcula que en la actualidad se ingresan a los buscadores más de 40 mil búsquedas por segundo y que -para el año 2020- las personas produciríamos casi 2Mb de información por segundo. Y estos datos, tal como sucede ahora, provendrán de las más diversas fuentes. Las redes sociales, los datos públicos, la información privada y la industrial e, incluso, las ciencias que producen grandes volúmenes de datos y que requieren gran capacidad de cómputo son algunos ejemplos que reflejan la heterogeneidad de las formas de producción, formatos, cantidad y calidad en que surgen los datos.

Esta nueva realidad -el crecimiento exponencial del universo digital- conlleva desafíos urgentes que resolver y es el objetivo con el que se crea

el Instituto Milenio Fundamentos de los Datos, financiado por la Iniciativa Científica Milenio.

LOS DATOS DESDE SUS FUNDAMENTOS

El Instituto Milenio Fundamentos de los Datos se encuentra albergado en la Universidad de Chile y la Pontificia Universidad Católica de Chile, con investigadores de ambas casas de estudios, además de académicos pertenecientes a la Universidad Técnica Federico Santa María, la Universidad de Concepción, la Universidad de Talca y la Universidad Diego Portales. Asimismo, este centro cuenta con una red de colaboración investigativa con otras



universidades del país y del extranjero en todas las áreas de su quehacer, con la meta de hacer investigación de frontera en ciencia de los datos.

“La información que produce la sociedad solo puede ir en aumento”, explica Pablo Barceló, Profesor Titular del Departamento de Ciencias de la Computación de la Universidad de Chile y director alterno del Instituto. “Se estima a nivel general que el universo digital, es decir, el flujo y cantidad de bytes que la sociedad crea y reproduce y que cada vez representa más a la realidad análoga, pasará de 4,4 zetta-bytes en 2013 a 44ZB en 2020: esto equivale a 44 billones de GB, con una velocidad de crecimiento que al menos se duplica cada dos años”, señala Barceló.

POLÍTICA Y CIENCIA DE DATOS PARA EL DESARROLLO DE LA SOCIEDAD

Ante esta realidad, Fundamentos de los Datos busca convertirse en un referente en ciencia de los datos, investigando y comprendiendo a esta unidad básica de la misma forma que la biología estudia a la célula o la física al átomo: como una unidad y a la vez como parte de sistemas y dinámicas mucho más complejos.

La gran riqueza del Instituto Fundamentos de los Datos radica, precisamente, en el encuentro de disciplinas tradicionalmente vistas como distantes entre sí. Es así como desde las cien-

cias de la computación, las matemáticas, estadística, ciencia política y comunicaciones, se abordará al dato desde sus orígenes, pasando por sus formatos y plataformas de distribución, hasta los usos e impactos que tienen en la sociedad.

Este enfoque es el que se refleja también en los cinco proyectos emblemáticos definidos por el Instituto:

1. Acercamiento sectorial en poblaciones de Santiago y la Araucanía

Esta iniciativa busca integrar y visualizar datos de sectores de Chile con conflictos sociales importantes, como demografía, salud, educación, seguridad, características urbanas y accesibilidad, entre otros elementos, en Santiago

y la Araucanía, para idealmente generar análisis de alto valor social que pueda ser puesto a disposición de las comunidades con miras a su estudio y potencial mejoramiento social.

2. Desarrollo e implementación de un lenguaje de consulta estándar para grafos

En la actualidad una de las formas más utilizadas para el almacenamiento de información son las bases de datos de grafos, en las que los datos generan nodos y conexiones entre sí, como sucede en forma casi natural en las redes sociales. Los sistemas basados en grafos tienen la ventaja de que entregan flexibilidad a un modelo, ya que permiten integrar datos de manera sencilla a la dinámica ya configurada. Hoy existen iniciativas para desarrollar lenguajes estándares de consulta de grafos y en dos de ellas, de carácter internacional, participan investigadores del Instituto: G-Core y GraphQL. Este proyecto busca desarrollar técnicas que permitan implementar estos lenguajes de forma eficiente y correcta.

3. Lenguajes de consulta con capacidades estructurales y analíticas

Extraer información relevante de distintas fuentes de datos es una de las grandes tareas en ciencias de la computación. Para lograr las funcionalidades analíticas que permitirían alcanzar este objetivo se utiliza hoy álgebra lineal en combinación con álgebra relacional, lo que da origen a lenguajes poderosos en términos de su capacidad de extracción de información relevante, pero que son computacionalmente complejos de evaluar. El desafío que este proyecto abordará es cómo equilibrar las funcionalidades de ambos modelos para generar lenguajes de consulta que sean más manejables en términos de su complejidad computacional, más fáciles de implementar y que permitan una mejor comprensión de su semántica y poder expresivo.

4. Inteligencia Artificial (IA) con explicación

Los sistemas de inteligencia artificial (IA) basados en deep learning han mostrado gran

éxito en el último tiempo. Sin embargo, una desventaja de este método es su operación tipo “caja negra”, es decir, es complejo precisar las motivaciones detrás de cada inferencia hecha por un sistema de IA. Comprender el tipo de información y los procesos que usan estos modelos en su operación es un tema de máxima relevancia en el área. Por una parte, tiene componentes prácticos y éticos que están comenzando a ser normados. Pero hay problemas aún más fundamentales: la creación de nuevas técnicas que permitan guiar el proceso de entrenamiento de una IA de manera que no memoricen patrones, sino que generen representaciones a través de meta-procesos de aprendizaje. Esto puede tener gran impacto para entender y mejorar la operación y uso de estos modelos.

5. Estudio y desarrollo de estructuras de información robustas para el fortalecimiento de la democracia

En el contexto actual en donde se ha hecho evidente la existencia de un “desorden informativo”, las fuentes tradicionales de producción y transmisión de información han tendido a perder la confianza de la sociedad y, por ende, ha disminuido su relevancia para la opinión pública. Se hace por lo mismo cada vez más importante estudiar y facilitar la generación de información robusta a partir de las nuevas fuentes de información, que van desde la Web, pasando por las redes sociales y el contenido multimedia, al texto expresado en lenguaje natural, entre otras.

Este proyecto se ha planteado comprender las dificultades y desafíos que implica generar estructuras de información robustas, para desarrollar una suerte de antídoto a la desinformación o información de carácter malicioso, priorizando esta investigación en los temas que pueden afectar el correcto funcionamiento de la democracia, como el déficit de legitimidad, la consolidación de filtros-burbuja que radicalizan las opiniones y la capacidad de movilización negativa de las *fake news*, entre otras. ■

Investigadores seniors del Instituto Milenio Fundamentos de los Datos:



Claudio Gutiérrez

Doctor en Ciencias de la Computación, Universidad de Wesleyan, Connecticut, Estados Unidos; académico de la Universidad de Chile.



Leopoldo Bertossi

Doctor en Matemáticas, Pontificia Universidad Católica de Chile; académico de la Pontificia Universidad Católica de Chile.



Ricardo Baeza-Yates

Doctor en Ciencias de la Computación, Universidad de Waterloo, Canadá; académico de la Universidad de Chile.

Investigadores asociados:



Marcelo Arenas
 Doctor en Ciencias de la Computación, Universidad de Toronto, Canadá; académico de la Pontificia Universidad Católica de Chile.



Pablo Barceló
 Doctor en Ciencias de la Computación, Universidad de Toronto, Canadá; académico de la Universidad de Chile.



Juan Reutter
 Doctor en Informática, Universidad de Edimburgo, Reino Unido; académico de la Pontificia Universidad Católica de Chile.



Jorge Pérez
 Doctor en Ciencias de la Ingeniería, Pontificia Universidad Católica de Chile; académico de la Universidad de Chile.



Bárbara Poblete
 Doctora en Ciencias de la Computación y Comunicación Digital, Universitat Pompeu Fabra, Barcelona, España; académica de la Universidad de Chile.



Éric Tanter
 Doctor en Ciencias de la Computación, Universidad de Chile y de la Universidad de Nantes, Francia; académico de la Universidad de Chile.



Benjamín Bustos
 Doctor en Ciencias Naturales, Universidad de Constanza, Alemania; académico de la Universidad de Chile.



Andrea Rodríguez
 Doctora en Ciencia e Ingeniería de la Información Espacial, Universidad de Maine, Estados Unidos; académica de la Universidad de Concepción.



Gonzalo Navarro
 Doctor en Ciencias mención Computación, Universidad de Chile; académico, Universidad de Chile.



Susana Eyheramendy
 Doctora en Estadística, Universidad de Rutgers, Estados Unidos; académica de la Pontificia Universidad Católica de Chile.



Marcelo Mendoza
 Doctor en Ciencias mención Computación, Universidad de Chile; académico de la Universidad Técnica Federico Santa María.



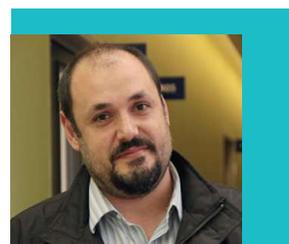
Sebastián Valenzuela
 Doctor en Comunicaciones, Universidad de Texas en Austin, Estados Unidos; académico de la Pontificia Universidad Católica de Chile.



Álvaro Soto
 Doctor en Ciencias de la Computación, Universidad de Carnegie Mellon, Estados Unidos; académico de la Pontificia Universidad Católica de Chile.



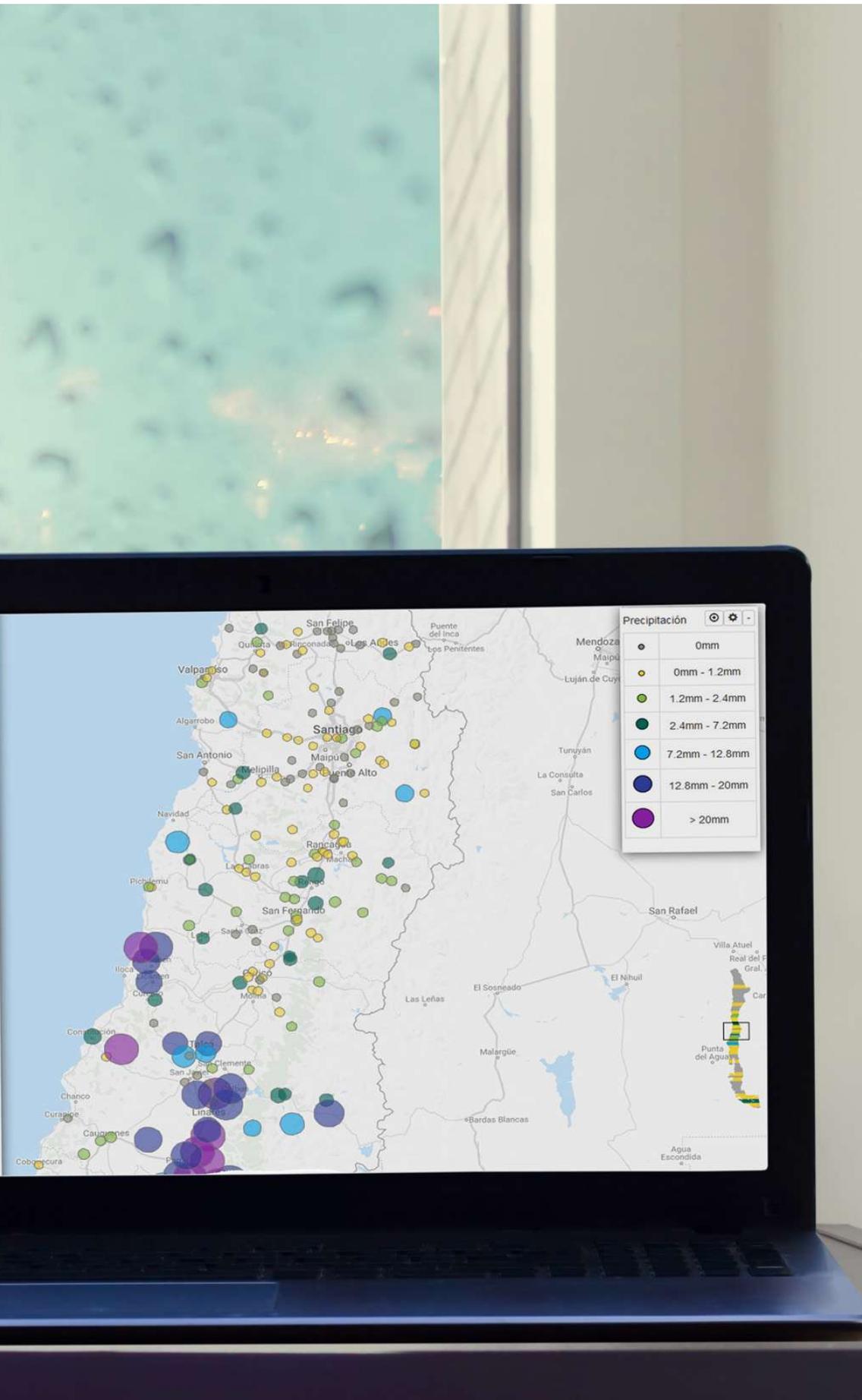
Aidan Hogan
 Doctor en Ciencias de la Computación, Universidad Nacional de Irlanda; académico de la Universidad de Chile.



Juan Pablo Luna
 Doctor en Ciencia Política, Universidad de Carolina del Norte en Chapel Hill, Estados Unidos; académico de la Pontificia Universidad Católica de Chile.

¿DÓNDE ESTÁ LLOVIENDO AHORA? PREGÚNTALE A VISMET





MARÍA CECILIA BASTARRICA

Profesora Asociada del Departamento de Ciencias de la Computación de la Universidad de Chile. Académica a cargo del curso Proyecto de Software. PhD Computer Science and Engineering, University of Connecticut.

Líneas de investigación: ingeniería de software, líneas de productos de software, mejora de procesos de software, desarrollo de software dirigido por modelos. cecilia@dcc.uchile.cl



RENÉ GARREAUD

Profesor Titular del Departamento de Geofísica de la Universidad de Chile. Subdirector del Centro de Excelencia FONDAP de Investigación del Clima y la Resiliencia (CR2). Doctor en Meteorología, Universidad de Washington, Estados Unidos.

Sus líneas de trabajo incluyen la dinámica del clima, cambio climático y meteorología sinóptica, con énfasis en América del Sur. rgarreau@dgf.uchile.cl



FIGURA 1. VISTA DE DOS ESTACIONES METEOROLÓGICAS AUTOMÁTICAS (EMAs) SOBRE LA TERRAZA DEL DEPARTAMENTO DE GEOFÍSICA.

La meteorología es una ciencia que tiene dos pilares. Uno de ellos es el conjunto de leyes físicas de la atmósfera y otras componentes del sistema terrestre. Su integración numérica requiere grandes capacidades computacionales y permite el pronóstico del tiempo ("*forecasting*") hasta dos semanas en el futuro. El otro pilar es el mundo de las observaciones, sin las cuales no se podrían inicializar los modelos, contrastar sus resultados y hacer sentido de múltiples fenómenos atmosféricos. Más aún, el pronóstico meteorológico de muy corto plazo (0-6 horas) se basa mayormente en un tratamiento estadístico de las condiciones observadas actuales y pasadas. El así llamado "*nowcasting*" es de especial valor en condiciones extremas; por ejemplo, para determinar si ocurrirá una crecida súbita en algún cajón cordille-

rano necesitamos saber qué está pasando ahora en la cordillera.

Existen múltiples fuentes de datos meteorológicos, incluyendo mediciones desde satélites y otros sensores remotos, pero un conjunto esencial son las observaciones directas sobre la superficie del planeta, en especial de temperatura, precipitación, viento y presión. La observación sistemática, registro y almacenamiento de las condiciones de la atmósfera son realizados tradicionalmente por los servicios nacionales de meteorología e hidrología, pero en la actualidad se han involucrado muchas empresas privadas (como las mineras en Chile) e instituciones académicas. A lo anterior ha contribuido la disponibilidad de estaciones meteorológicas automáticas

(EMAs, Figura 1), que a un costo relativamente bajo permiten mediciones con alta resolución temporal (por ejemplo, cada 15 minutos), autónomas (la mayoría de las EMAs se alimenta con energía solar) y con la posibilidad de transmitir sus datos a servidores remotos mediante telefonía celular, radio o enlace satelital.

Aunque en Chile no existe un catastro completo de las EMAs, estimamos que al menos hay 1.000 estaciones operativas a lo largo del territorio nacional. Un número importante de ellas transmite su data en tiempo real, de las cuales más de 400 pertenecen a la Dirección General de Aguas (DGA) del Ministerio del Obras Públicas, un número similar a la red AgroMET, un consorcio público-privado

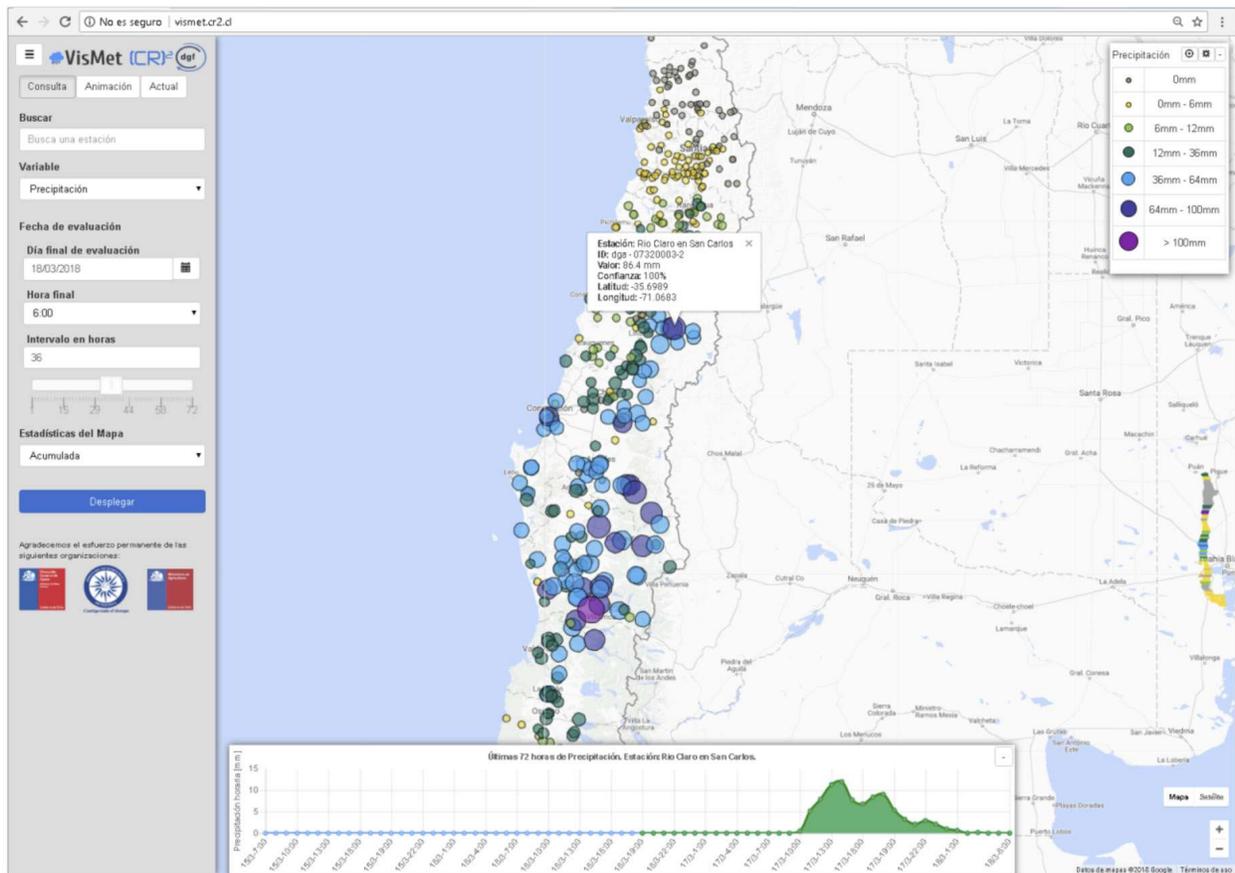


FIGURA 2. CAPTURA DE PANTALLA DE LA VERSIÓN ORIGINAL DE VISMET (ENERO 2018).

organizado por el Ministerio de Agricultura, y cerca de 80 a la Dirección Meteorológica de Chile (DMC). El Servicio Meteorológico de la Armada de Chile también dispone de cerca de 40 EMAs a lo largo de la costa. Todos estos servicios cumplen una labor formidable velando por la calidad y continuidad de esas estaciones, recopilando sus datos y haciéndolos disponibles en sus páginas web. A pesar del gran número de EMAs en Chile la cobertura es insuficiente, especialmente en la cordillera y regiones extremas.

El esquema general seguido por las instituciones operativas es permitir el acceso a las series de tiempo de cada una de sus estaciones. Eso dificulta responder algunas preguntas básicas, como: ¿Dónde está lloviendo en este

preciso momento? ¿Qué zona afectó la helada del 27 de julio? ¿Cuál fue la intensidad de una ola de calor el 31 de enero? Responder estas preguntas es el punto de partida para el diagnóstico de una serie de procesos y fenómenos ambientales, como sería una inundación, daños a la agricultura e incendios forestales, respectivamente.

Motivados por esta dispersión de los datos meteorológicos en Chile, que dificulta su puesta en valor y generación de información ambiental, durante el año 2017 desarrollamos VisMet, Visualizador Meteorológico. La idea conceptual es muy clara: visualizar en un mapa geográfico los datos de precipitación y temperatura horaria actual o de algún momento pasado, registradas por las EMAs a lo largo

de Chile. Nuestra experiencia con plataformas similares en otros países y colaboración en el desarrollo del Explorador Climático, una herramienta análoga pero enfocada en datos mensuales o anuales, aportó condiciones adicionales para el desarrollo de VisMet: simplicidad e inmediatez de la plataforma web.

Todo claro en esa visión, pero el diablo está en los detalles... no es simple construir un sistema robusto y 24/7 de "antenaje" de datos vía web, su integración y homologación (no todos los datos están disponibles en forma simultánea) y su correcta visualización. El desafío fue planteado a un grupo de estudiantes¹ del Departamento de Ciencias de la Computación (DCC) que participaban en el

1. Los alumnos del DCC que realizaron el desarrollo de VisMet fueron: Nicolás Caracci, Andrés Ferrada, Maximiliano Kauer, Belisario Panay, Juan Pablo Paulsen y Sergio Peñafiel. En este proceso contaron con el apoyo de Francisca Muñoz (CR2) y Felipe Saavedra (DGF).



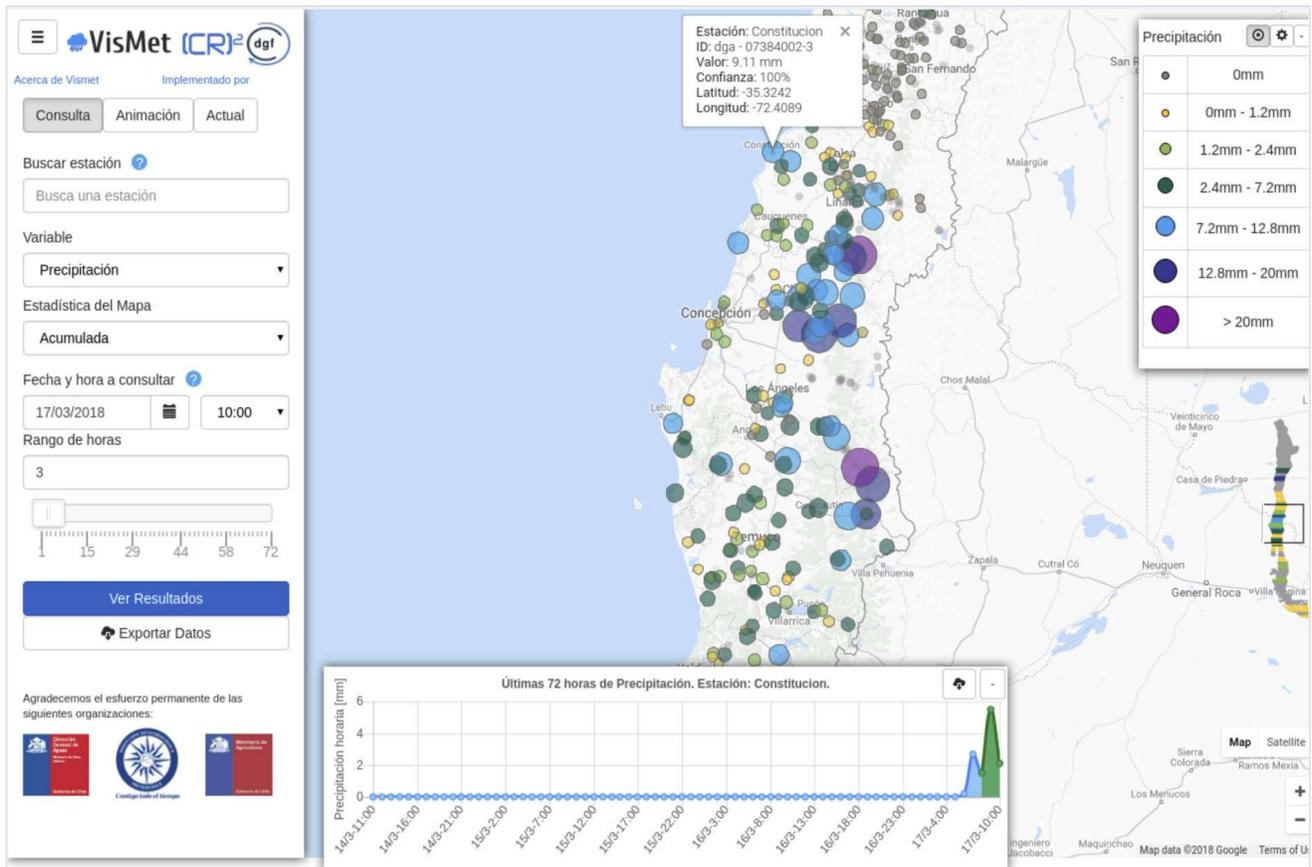


FIGURA 3. CAPTURA DE PANTALLA DE LA VERSIÓN MEJORADA DE VISMET (AGOSTO 2018).

curso Proyecto de Software en el semestre de primavera 2017. El desarrollo también contó con el apoyo del Centro del Clima y la Resiliencia (CR2, Francisca Muñoz) y el Departamento de Geofísica de la Universidad de Chile (DGF, Felipe Saavedra).

El curso Proyecto de Software (CC5402) es el último obligatorio de Ingeniería Civil en Computación y se desarrolla en el décimo semestre. Este curso es parte de la malla de la especialidad desde hace casi veinte años, pero ha experimentado cambios metodológicos periódicos. Desde hace ya algunos años los alumnos desarrollan sistemas de software reales en equipos de cinco a siete alumnos trabajando en las instalaciones del cliente. Allí tienen constante interacción con los clientes y usuarios

finales de modo de asegurar que el software desarrollado efectivamente esté alineado con las necesidades. A pesar de que se permite que cada equipo gestione su proyecto como crea más apropiado, el curso sugiere la aplicación de prácticas de desarrollo ágil, en particular cliente *in situ*, desarrollo iterativo, y desarrollo dirigido por valor, entre otras.

La página web de VisMet fue hecha pública a comienzos de 2018 (Figura 2) montada en los servidores del CR2 (<http://vismet.cr2.cl/>). Al iniciar una sesión se despliega en el mapa central la precipitación acumulada en las últimas 3 horas -una especie de fotografía de lo que está ocurriendo en el momento- en un conjunto de más de 700 estaciones a lo largo de Chile operadas por la DMC, DGA y AgroMet. El

usuario puede elegir otra variable (precipitación o temperatura), periodo de análisis (entre 1 y 72 horas), tiempo final de la consulta (cualquier día/hora de los últimos 10 años) y la estadística a desplegar (acumulación, promedio, valor máximo o mínimo). Sobre el mapa se puede controlar la región de visualización (Zoom in - Zoom out), la escala de colores y obtener la serie de valores horarios de las estaciones seleccionadas. En julio de 2018 se actualizó ligeramente VisMet, reordenando el menú de selección y agregando la funcionalidad para descargar los datos del mapa y series de tiempo (Figura 3).

La divulgación de VisMet se ha realizado a través de la lista de contacto del CR2 y el DGF, con énfasis en investigadores y profesionales

del ámbito de la meteorología e hidrología. Durante los meses de verano de 2017, cuando el tiempo es muy estable, hubo escasas visitas a VisMet, pero con la llegada de las primeras lluvias a Santiago las visitas crecieron sustancialmente (**Figura 4**), llegando a más de 40.000 al momento de escribir esta nota. Hay una gran cantidad de visitantes los días de lluvia, pero hay un nivel basal (~50) debido al uso rutinario de VisMet en instituciones operativas las cuales se han mostrado muy satisfechas por la integración multiagencia que esta plataforma ha logrado. VisMet es empleado regularmente en instituciones que enfrentan emergencias (ONE-MI, CONAF) cuando la situación lo requiere y

por investigadores quienes encuentran aquí por primera vez una base de datos meteorológica consolidada a nivel horario, abriendo nuevas ventanas al entendimiento de eventos extremos. El creciente uso de VisMet con fines operativos, profesionales, científicos o simplemente para saber dónde está lloviendo ahora, es un motivo de satisfacción para el equipo tras su desarrollo, y a la vez una motivación para mantener operativa y eventualmente expandir esta simple pero potente plataforma en tiempo real.

En el curso Proyecto de Software se abordan proyectos de organizaciones de variada natu-

raleza: reparticiones de la universidad, centros de investigación, start-ups, organizaciones públicas, fundaciones o empresas privadas. La selección de proyectos se realiza de acuerdo con el potencial impacto del producto desarrollado, la factibilidad o riesgo que implica el proyecto, y el aprendizaje técnico que puedan obtener los alumnos. La contraparte técnica del proyecto VisMet quedó gratamente satisfecha tanto por la calidad del producto final, como por el compromiso, dedicación y flexibilidad de los alumnos del DCC que desarrollaron esta plataforma. Estos resultados son similares a los obtenidos en otros proyectos. ■

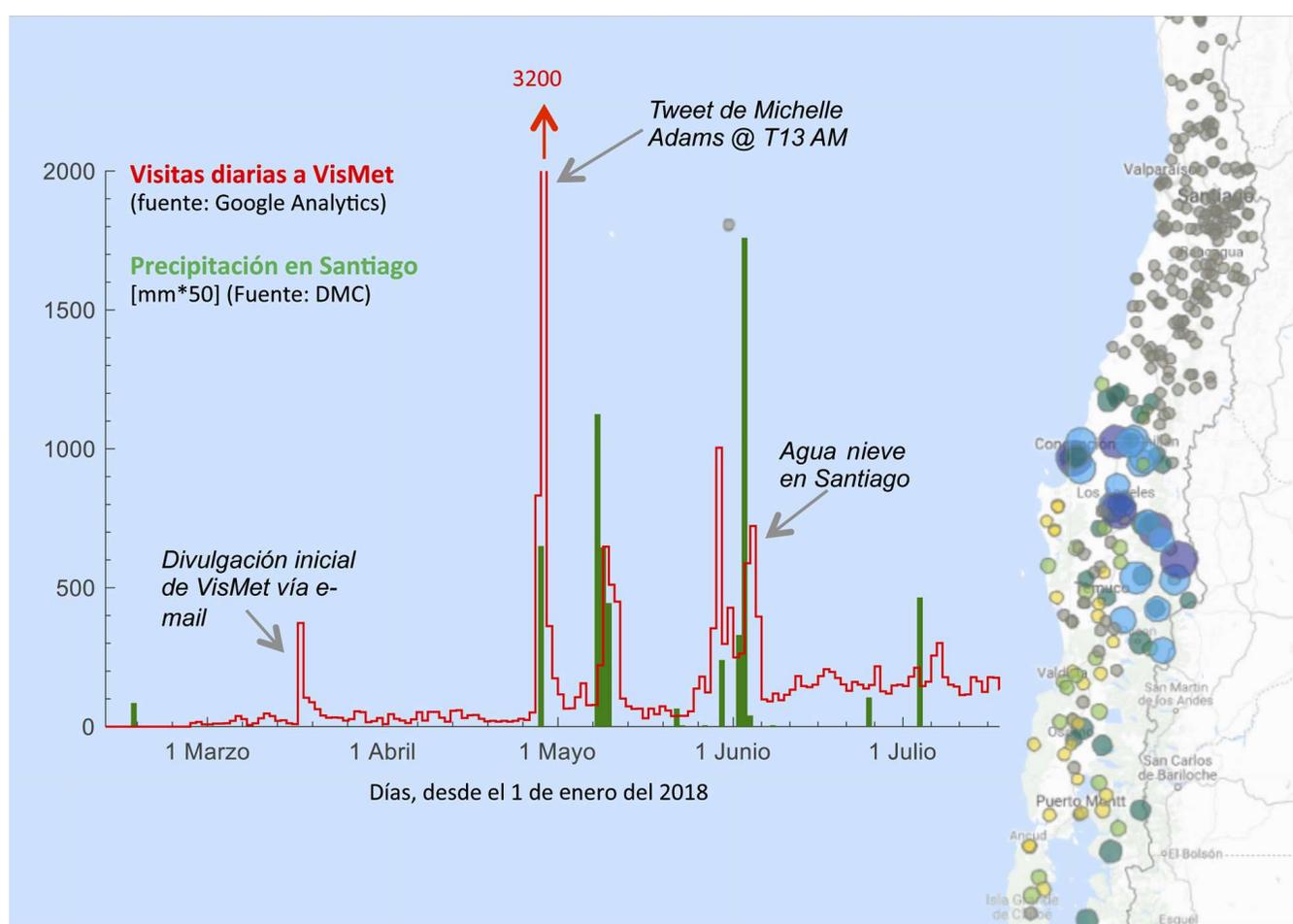


FIGURA 4. NUMERO DE VISTAS DIARIAS A VISMET ENTRE MARZO Y AGOSTO DE 2018 (LÍNEA ROJA) SOBREPUESTA SOBRE LA PRECIPITACIÓN DIARIA EN SANTIAGO (BARRAS VERDES). EL MAPA DE FONDO ES LA PANTALLA DE VISMET.

EL TIEMPO DIRÁ...

```
elif _operation == "MIRROR"  
    mirror_mod.use_x  
    mirror_mod.use_y  
    mirror_mod.use_z  
elif _operation == "MODIFY"  
    mirror_mod.use_x  
    mirror_mod.use_y  
    mirror_mod.use_z  
  
#selection at the  
mirror_ob.select= 1  
modifier_ob.select=1  
bpy.context.scene.obj  
print("Selected" + st  
#mirror_ob.select  
bpy.context.se
```

En noviembre de 2017 y abril de 2018, dos artículos de mi autoría recibieron premios por su impacto destacado en los últimos diez años. El primero fue publicado en el Dynamic Languages Symposium (DLS) del año 2007, y el segundo en la International Conference on Aspect-Oriented Software Development (AOSD) del año 2008.

Por su cercanía temporal, el Editor de la Revista Bits me pidió escribir una nota sobre ambos premios, quizás anticipando alguna relación directa entre estas investigaciones. Sin embargo, me asombra observar que ambos hitos ¡no guardan ninguna relación entre sí! Antes de detallar sus diferencias, voy a resumir brevemente los aportes científicos de cada uno de estos artículos.

El artículo publicado en DLS 2007 se titula **“Mirages: Behavioral Intercession in a Mirror-based Architecture”**. Este trabajo presenta una forma estructurada de hacer “reflexión de comportamiento” en programas, es decir, que un programa pueda acceder y modificar ciertos aspectos de su ejecución de manera dinámica, sin comprometer aspectos de seguridad y modularidad. Anteriormente, en 2004, Bracha y Ungar (ambos en Sun Microsystems en ese entonces) habían popularizado las arquitecturas de espejos (“mirrors”) para proveer reflexión en un lenguaje de programación, respetando principios de diseño fundamentales como encapsulación y estratificación. La contribución técnica del artículo fue mostrar cómo extender dichas arquitecturas para proveer una forma de reflexión que la propuesta original de Bracha y Ungar no soportaba. Si bien parece medio esotérico, el impacto destacable de este trabajo radica en que el mecanismo descrito tuvo, años después, una fuerte influencia en el diseño de los “proxies” del lenguaje JavaScript, que hoy en día es el lenguaje de programación más usado para desarrollos web.

El artículo publicado en AOSD 2008 se titula **“Expressive Scoping of Dynamically-Deployed Aspects”**. El artículo propone un modelo muy expresivo para aumentar la flexibilidad de los lenguajes de programación por aspectos a través de un manejo avanzado de la propagación de aspectos durante la ejecución de un programa. En breves palabras, la programación por aspectos permite definir de manera modular las preocupaciones transversales de una aplicación, como lo son, por

ejemplo, la seguridad y la resiliencia a fallas. En ese entonces era un tema muy activo. En el paper, se proponen estrategias de control de alcance (*scoping*) que aumentaron significativamente el poder de los aspectos de una manera particularmente útil para ciertas aplicaciones, como la seguridad en general, y el control de acceso en particular. En trabajos futuros junto a otros colegas investigamos el manejo de seguridad con aspectos, para lo cual el mecanismo expresivo de *scoping* fue una piedra esencial.

Ahora volviendo a la relación entre ambos premios, parece ser que lo único que los acerca es la fecha. Científicamente se ubican en contextos muy distintos, como fue resumido anteriormente: uno trata de reflexión (en un lenguaje orientado a objetos), y el otro trata de aspectos (en un lenguaje funcional).

Además, el primero (DLS) fue fruto de una colaboración con varios colegas (Stijn Mostinckx, Tom Van Cutsem y Stijn Timbermont, de la Vrije Universiteit Brussel en Bélgica). Mi participación fue más bien secundaria, siendo ellos los más proactivos (Tom Van Cutsem es quien, luego de terminar su doctorado sobre el tema, empezó a trabajar en el esfuerzo ECMAScript, el estándar detrás de JavaScript). En cambio, el segundo (AOSD) lo escribí como único autor, fruto de mucho esfuerzo y transpiración, intentando publicar por primera vez en esa conferencia tan prestigiosa que hasta ese entonces me había negado el privilegio de aceptar mis trabajos.

El artículo DLS era una continuación bastante directa de mi tesis de doctorado, mientras que el artículo AOSD era el inicio de una nueva línea de investigación. De hecho, en los años siguientes, abandoné el hilo del primer artículo, mientras el segundo se transformó en la base de muchos trabajos, incluyendo la tesis de doctorado de Rodolfo Toledo sobre aspectos y seguridad.

Finalmente, el artículo DLS tuvo un impacto industrial real (¡y totalmente no anticipado!), sin atraer muchas citas por parte de otros trabajos académicos. Para el artículo AOSD, la situación fue exactamente inversa: tuvo impacto académico, con muchas citas y continuaciones científicas -incluso de otros investigadores- pero no tuvo ningún impacto industrial.

¿Quién dijo que la academia es monótona? ■



**ÉRIC
TANTER**

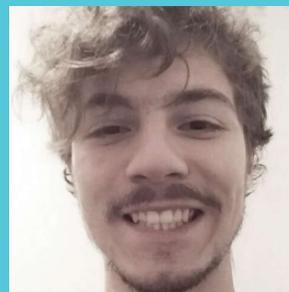
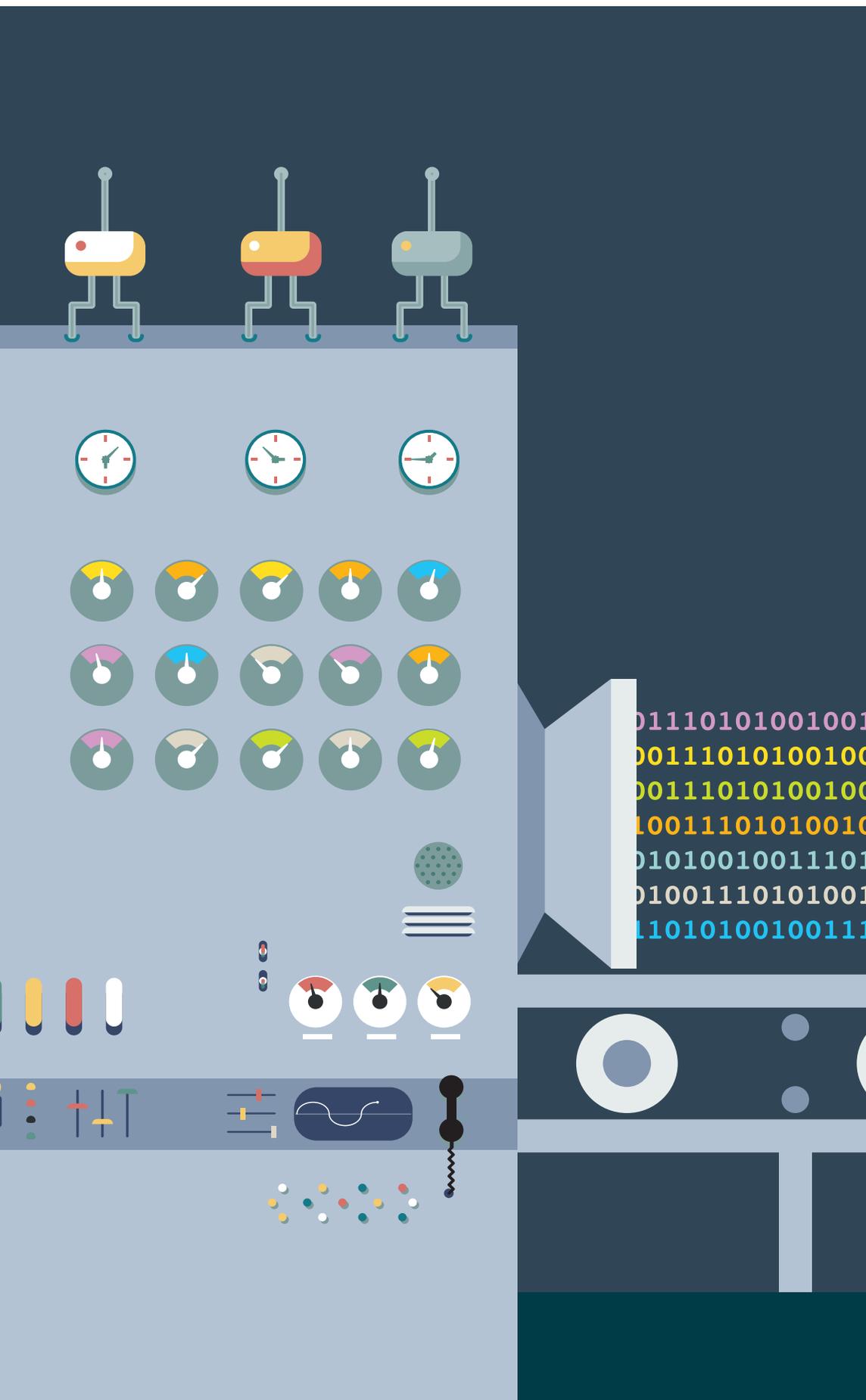
Profesor Titular del Departamento de Ciencias de la Computación de la Universidad de Chile, e Investigador Asociado del Instituto Milenio Fundamentos de los Datos. Ph.D. Computer Science, Universidad de Nantes, Francia y Universidad de Chile (2004).

Líneas de investigación: lenguajes de programación, ingeniería de software.

etanter@dcc.uchile.cl

The background features a dark blue field with stylized server racks on the right and a teal text box on the left. Curved lines of binary code (0s and 1s) in yellow, purple, and light blue are scattered across the top and middle sections.

NAVEGANDO SOBRE CORRIENTES DE DATOS Y CÓMO NO AHOGARSE EN EL INTENTO



ALEJANDRO GREZ

Alumno de Doctorado en Ciencias de la Computación de la Pontificia Universidad Católica de Chile. Ingeniero Civil en Computación y Magíster en Ciencias de la Ingeniería, Pontificia Universidad Católica de Chile.

Líneas de investigación: manejo de datos streaming, complex event processing, teoría de autómatas, lógica computacional.
ajgrez@uc.cl



CRISTIÁN RIVEROS

Profesor Asistente del Departamento de Ciencia de la Computación de la Pontificia Universidad Católica de Chile. Investigador del Instituto Milenio Fundamentos de los Datos. DPhil in Computer Science, Oxford University.

Líneas de investigación: manejo de datos, sistema de base de datos, datos streaming, extracción de información, fundamentos de la computación, modelos de computación.
cristian.riveros@uc.cl

MANEJO DE DATOS STREAMING

Redes de sensores, redes sociales, tráfico en Internet, mediciones meteorológicas, observaciones astronómicas, son algunos escenarios en donde se generan gigabytes de datos continuamente y a gran velocidad. Pero, ¿de cuánta velocidad estamos hablando exactamente? Actualmente, un auto promedio cuenta con aproximadamente 100 sensores que miden datos relevantes para su funcionamiento, como el estado del motor o el nivel de bencina. A través de estos sensores, un auto genera alrededor de 10 gigabytes de datos cada hora. Otro ejemplo más cotidiano es el caso de Twitter. Si juntamos todos los tweets generados por los usuarios de esta red social, a cada segundo se twitean alrededor de 6 mil tweets, lo que equivale a 20 gigabytes generados por cada hora. Estos escenarios, conocidos como streams de datos, son ejemplos del fenómeno del Big Data y, en particular, aluden a la dimensión de velocidad: el problema de ser capaces de procesar grandes cantidades de datos en tiempo real.

El gran desafío que plantean estos escenarios es ¿cómo procesar de manera eficiente estos datos para obtener el mayor valor posible de ellos? Para responder esta pregunta, primero debemos notar que en muchos casos no es factible ni siquiera almacenar todos estos datos para procesarlos más tarde. En el caso de un auto, en un día habrá generado más de 240 gigabytes de datos, superando por lejos sus capacidades de almacenamiento. Lo segundo que debemos notar es que, incluso si logramos almacenar esta cantidad de datos, no tendremos quizás el tiempo suficiente para procesarlos. Hay que pensar que solo leer 240 gigabytes, y sin hacer ningún cálculo, nos tardaría alrededor de dos minutos en un procesador convencional. Por último, los datos generados por Twitter o en un auto son de importancia en el segundo que se generan, desvalorizándose rápidamente a medida que avanza el tiempo. En otras palabras, queremos saber sobre un “trending topic” de nuestro interés en Twitter, o de una posible falla del auto apenas suceda.

¿Cómo, entonces, procesar de manera eficiente estos streams de datos? Los algoritmos stream-

ing aparecen a principios de este siglo como una respuesta para obtener información valiosa de estos grandes flujos de información. Dado que no tenemos ni tiempo ni espacio para acordarnos del pasado, un algoritmo streaming busca mantener una representación muy pequeña de los datos que ha visto (el pasado) que contenga solo la información que es de interés. Cada vez que recibe un dato nuevo, el algoritmo streaming actualiza su representación compacta con el dato (sin aumentar su tamaño considerablemente) y tomando tiempo que es proporcional solo a este nuevo dato. En otras palabras, no importa cuántos datos ha leído en el pasado, el tiempo de actualización por cada dato nuevo será el mismo. Para ilustrar la idea de un algoritmo streaming, suponga que usted está recibiendo el siguiente stream, en donde cada evento es un número:

↓
3, 4, 8, 10, 12, 34, 12, 54, 65, 23, ...

La condición es que, en cualquier momento, usted debe ser capaz de entregar el promedio aritmético de todos los números que han llegado hasta ese momento. Una posibilidad es que usted almacene toda la secuencia que ha recibido, y cuando le pregunten por el promedio, sume todos los números anteriores y divida por la cantidad de números. Por ejemplo, si estamos en el momento indicado por la flecha y se nos pregunta por el promedio, podemos sumar los números y dividir por la cantidad de números (9 hasta el 65) dando:

$$\frac{(3 + 4 + 8 + 10 + 12 + 34 + 12 + 54 + 65)}{9} = 22.44$$

Esto cumple con entregar el promedio de los números; sin embargo, es muy ineficiente ya que tengo que almacenar todos los números anteriores (usando mucho espacio) y sumarlos todos cada vez que necesito el promedio (mucho tiempo). En cambio, un algoritmo streaming muy sencillo puede hacer esto de una manera muy eficiente almacenando la suma parcial de los datos que hemos visto, **sum**, junto con la cantidad de números, **cant**. Luego, para cada número nuevo n actualiza los valores como:

```
sum := sum + n
cant := cant + 1
```

Ahora, cada vez que se nos solicite el promedio de los números anteriores, solo necesitamos hacer el cálculo **sum/cant**, entregando el promedio esperado. Note que ahora lo único que debemos almacenar son estos dos números, y que actualizar estos valores solo nos toma dos operaciones por cada nuevo número.

Lamentablemente, no todos los problemas en procesamiento de datos streaming tienen una solución tan sencilla como la anterior. En esta área se han estudiado diversos problemas de procesamiento streaming como, por ejemplo, encontrar los datos más frecuentes o contar la cantidad de datos distintos que han ocurrido en un stream. Estos problemas muchas veces no tienen un algoritmo que pueda calcular, con poco espacio y tiempo, el resultado exacto. Por esto, muchos algoritmos streaming usan aleatoriedad para calcular un valor aproximado a la solución, cambiando eficiencia por precisión. Ésta es una área muy desafiante del diseño de algoritmos que ha logrado grandes avances en los últimos años.

SISTEMAS DE BASES DE DATOS STREAMING

A simple vista, los algoritmos streaming parecen ser una buena herramienta a la hora de manejar grandes volúmenes de datos, pero en la práctica resultan demasiado difíciles de utilizar o de programar. Por otro lado, tampoco queremos tener que programar un algoritmo desde cero cada vez que estamos interesados en extraer cierta información valiosa de un stream. Para hacer esto, el uso de un sistema de manejo de datos streaming (similar a lo que se conoce como un sistema de bases de datos) parece el correcto: tener un sistema (software) que se encargue de leer y gestionar nuestro stream de datos, mientras nosotros solo debemos escribir nuestra pregunta con un lenguaje declarativo (como SQL) y luego dejar que el sistema se encargue de responder la pregunta eficientemente. Ésta es la dirección que se ha tomado en el área de manejo de eventos complejos (CEP, del inglés Complex Event Processing). Un sistema de CEP considera al stream de datos simplemente como una secuencia de eventos, donde cada dato es visto como un

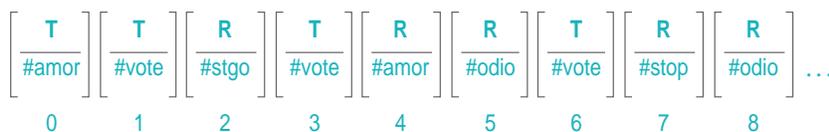


FIGURA 1
EJEMPLO DE UN STREAM DE DATOS EN TWITTER.

“Encuentra un tweet que contenga ‘#voteporTrump’ seguido de una respuesta que contenga ‘#odioaTrump’”

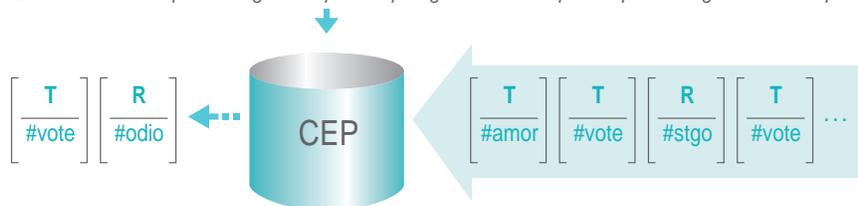


FIGURA 2
DIAGRAMA DEL FUNCIONAMIENTO DE UN SISTEMA DE MANEJO DE EVENTOS COMPLEJOS (CEP: COMPLEX EVENT PROCESSING).

evento o suceso. Por ejemplo, un tweet en Twitter puede ser visto como un evento $T_{(post)}$ donde T me dice que el evento es un tweet y $post$ es su contenido. Otro evento podría ser $R_{(reply)}$, donde la R indica que es un evento de respuesta, en este caso con un mensaje de respuesta $reply$ ¹. En la **Figura 1** podemos ver un posible stream de eventos T y R de Twitter, donde los eventos van llegando de izquierda a derecha (por presentación, los mensajes de cada tweet o respuesta han sido acortados a un hashtag). Ahora que tenemos un esquema que nos dice la estructura de los datos, podemos consultar al sistema CEP por una pregunta de interés o, en la nomenclatura de CEP, por un evento complejo. Situémonos en el caso que somos un periodista interesado en debates políticos y tenemos acceso al stream de eventos de tweets y respuestas de Twitter. Una pregunta de nuestro interés puede ser, por ejemplo, saber cuándo ocurre un debate en Twitter sobre Trump. Esta pregunta se podría enunciar de manera precisa como:

“ Encuentra un tweet que contenga ‘#voteporTrump’ seguido de una respuesta que contenga ‘#odioaTrump’ ”.

Si vemos nuestro stream de la **Figura 1**, una respuesta posible a esta pregunta es el par de eventos en las posiciones 1 y 5, que contienen

exactamente los datos que buscamos. En otras palabras, un evento complejo no es más que un conjunto de posiciones que certifican la ocurrencia del evento.

Formalizando las ideas anteriores, el funcionamiento de un sistema CEP se puede entender con el diagrama en la **Figura 2**. El sistema lee el stream de datos (llamados eventos) y los usuarios envían consultas de interés a la base de datos. Apenas el sistema verifica que este evento complejo ha ocurrido (por ejemplo llega una respuesta R con ‘#odioaTrump’ después de un tweet T con ‘#voteporTrump’), el sistema le entrega una notificación o alerta al usuario junto con los eventos que respaldan el evento complejo. Una vez que la alerta fue entregada al usuario, el sistema CEP sigue procesando los eventos en busca de nuevos eventos complejos que satisfagan la pregunta. Es importante notar aquí la diferencia de los sistemas CEP con un sistema de bases de datos tradicional (como Oracle o PostgreSQL). En un sistema tradicional los datos son estáticos (esto es, con pocas modificaciones) y son las consultas las que cambian continuamente. En cambio, en un sistema CEP las consultas son estáticas y son los datos los que cambian continuamente.

A pesar de que los sistemas CEP parecen una buena solución para gestionar datos streaming,

hasta el momento estos sistemas no han tenido el protagonismo que uno espera en la práctica. Actualmente existen algunos sistemas ya asentados en la industria (como EsperTech) y algunos prototipos en la academia que son usados como referentes. Lamentablemente, estos sistemas carecen de una base sólida, con lenguajes de consultas poco claros y técnicas de evaluación deficientes. De hecho, es mencionado continuamente en la literatura que los lenguajes de consultas para CEP carecen de una semántica clara, siendo muchas veces confusos y difíciles de entender. Nuestra experiencia también nos ha mostrado que muchos de estos sistemas son difíciles de utilizar y configurar, funcionando correctamente solo para grupos particulares de consultas.

En nuestro proyecto (FONDECYT 11150653) buscamos dar formalidad a estos sistemas CEP. Para ello, proponemos empezar desde cero, dando respuesta a las preguntas más básicas (¿qué es un evento?, ¿qué es un evento complejo?, etc.) y reconociendo cuáles son las componentes fundamentales de estos sistemas. Ya con esta base definida, desarrollamos algoritmos que aseguran una ejecución eficiente y eficaz del sistema. A continuación, presentamos una breve reseña de nuestra investigación junto con algunos de los resultados que hemos obtenido hasta el momento.

NUESTRA PROPUESTA

La primera tarea en esta investigación fue formalizar el lenguaje de consultas para definir los patrones que el usuario desea utilizar. En la literatura se han propuesto más de 20 lenguajes de consulta distintos, con variados operadores y semánticas. De todos estos operadores, reconocimos los más recurrentes y que son fundamentales para la detección de eventos complejos. Estos operadores incluyen el de secuenciación (;), selección, disyunción (or), e iteración (similar a una clausura de Kleene), entre otros. Para dar una idea general de nuestro lenguaje, veamos el siguiente ejemplo. Recuerden la pregunta del periodista que buscaba debates sobre Trump en el stream de tweets (T) y respuestas (R). Esta pregunta

1. Twitter maneja una estructura muchísimo más compleja de datos que incluye identificadores, fechas (timestamp), retweets, follow y unfollow, entre otros. Aquí usamos una visión más sencilla de los datos de Twitter para simplificar la presentación. →

se podría definir con nuestro lenguaje teórico² de la siguiente manera:

$(T; R)$ FILTER ($T.post = \#vote'$
AND $R.reply = \#odio'$)

Para simplificar la presentación, en la expresión anterior abreviamos '#voteporTrump' por '#vote' y '#odioaTrump' por '#odio'. Esta expresión se puede leer como: quiero un tweet **T** seguido de una respuesta **R**, donde el mensaje de **T** contiene '#voteporTrump' y el mensaje de **R** contiene '#odioaTrump'. Si preguntamos por esta consulta en nuestro sistema **CEP**, sobre el stream de la **Figura 1**, entregaría el par de datos en las posiciones 1 y 5, dado que cumplen con la especificación anterior. Aquí el operador de secuenciación (; en símbolos) permite que entre **T** y **R** haya una cantidad arbitraria de eventos. Éste es un comportamiento esperado de un sistema **CEP** debido a que los streams de datos usualmente contienen mucho ruido, y los eventos nunca suceden continuamente como uno esperaría. Parte de nuestra investigación consistió en dar un significado claro a cada uno de estos operadores, lo que resulta en que el usuario tenga claro lo que espera recibir en respuesta a consultas como la anterior.

Una vez que definimos y formalizamos nuestro lenguaje de consultas, el siguiente paso fue entender cómo evaluarlas. En general, para la evaluación de consultas en sistemas de bases de datos uno busca una representación "intermedia", esto es, un objeto abstracto que define la consulta pero que sea más fácil de entender, manejar y optimizar. Por ejemplo, en una base de datos relacional, toda consulta SQL hecha por un usuario es convertida por el sistema a una expresión en álgebra relacional, la cual es mucho más sencilla y fácil de trabajar computacionalmente. En nuestro caso, utilizamos un modelo de autómatas como representación interna, el cual llamamos "autómatas de eventos complejos" (CEA, del inglés Complex Event Automata). Un CEA es muy parecido a un autómata finito no-determinista, con la diferencia que (1) lee streams en vez de palabras, (2) sus transiciones son predicados en vez de letras, y (3) sus transiciones pueden producir



FIGURA 3
CEA QUE REPRESENTA LA PREGUNTA DE DEBATES EN TWITTER.

resultados (eventos complejos) además de leer eventos. Para tener una idea de lo que es un CEA, veamos el siguiente ejemplo. En la **Figura 3** podemos ver un CEA con tres estados, representados por círculos, y sus transiciones representadas por las flechas entre estados. La idea de estas flechas etiquetadas como ($P | \bullet$) o ($PI \circ$) es muy sencilla: si el dato d que leo cumple la condición P entonces cambia al siguiente estado y, si es de tipo \bullet , marca d como parte del evento complejo y, si no, omite d . Como ejemplo, la flecha que nos lleva del estado 1 al estado 2 representa que si el dato d es de tipo tweet **T** y su mensaje (**post**) contiene '#voteporTrump', entonces cambiamos al estado 2 y marcamos d como parte del output. Es fácil ver que el CEA de la **Figura 3** representa, de hecho, nuestra consulta de debates, pero ahora con un modelo abstracto y procedural. En nuestra investigación estudiamos CEA formalmente y lo proponemos como el corazón de la evaluación de consultas en CEP. Es importante mencionar que en la literatura ya se habían propuesto modelos de autómatas similares para evaluar este tipo de consultas; nuestra contribución fue formalizar estos modelos matemáticamente, entendiendo sus propiedades fundamentales.

Una vez que ya tenemos una manera de representar las consultas con nuestro modelo intermedio de autómatas, la pregunta es ¿cómo evaluamos ahora estas máquinas eficientemente? Un desafío intrínseco de los lenguajes de CEP, por su capacidad de omitir datos intermedios, es que muchos eventos complejos pueden satisfacer una consulta al mismo tiempo. Por ejemplo, para nuestra consulta de debates no solo 1 y 5 cumplen con las condiciones de un debate al llegar el evento 5, si no también 3 y 5. De hecho, es fácil ver que una respuesta '#odioaTrump' puede emparejar con varios

tweets de '#voteporTrump'. Por lo tanto, un sistema CEP debe llevar de manera streaming (o sea, con poco espacio y tiempo), varios potenciales eventos complejos durante su ejecución.

Una de las contribuciones más relevantes de nuestro trabajo fue definir formalmente garantías de eficiencia que los algoritmos streaming deben satisfacer para ser considerados "eficientes". Proponemos utilizar lo que llamamos "algoritmos de enumeración con demora constante" (en inglés *constant delay enumeration algorithms*). ¿Qué hacen estos algoritmos específicamente? En pocas palabras, estos algoritmos mantienen una representación compacta de todas las posibles respuestas parciales y, cada vez que un nuevo dato es recibido, actualizan esta representación compacta en tiempo proporcional al nuevo dato. Por otra parte, cuando en un momento cualquiera se requiere obtener las respuestas a la consulta, otro proceso independiente se encarga de construir todos los resultados desde la estructura compacta y enumerarlos, tomando un tiempo constante entre cada par de resultados entregados. Cabe destacar que ésta es la forma más eficiente posible de enumerar los resultados, ya que el tiempo que tarda depende solo del tamaño de la respuesta. La gracia de estos algoritmos es que separan el tiempo de actualización del sistema del tiempo de enumeración, siendo eficientes tanto en la evaluación como en la enumeración de los resultados.

En nuestro trabajo, diseñamos y proponemos algoritmos de enumeración con demora constante para un subgrupo importante de consultas en nuestro lenguaje. En particular, estos algoritmos aplican para consultas del tipo anterior, como nuestra consulta sobre debates políticos en Twitter. De hecho, estos algoritmos funcionan

2. Presentamos aquí una simplificación del lenguaje. En nuestro trabajo, también proponemos un lenguaje de usuario con sintaxis SQL.

órdenes de magnitud más rápido que los mejores sistemas desarrollados actualmente en la academia e industria. Para ver esta diferencia en la práctica, realizamos un experimento muy sencillo con las siguientes dos consultas:

$$Q_1 = A; B; C$$

$$Q_2 = A; B; C; D$$

La consulta Q_1 busca tres eventos de tipos A , B y C , en ese orden. Por otra parte, la consulta Q_2 además solicita que después del tercer evento C debe ocurrir un evento de tipo D . Con estas consultas realizamos un experimento que llamamos de "stress": cada consulta es evaluada sobre un stream donde todos los eventos son generados con distribución uniforme exceptuando el último, que ocurre solo al final (el último evento es C para Q_1 y D para Q_2). Esto implica que un sistema CEP no encontrará ningún

resultado hasta que el último evento sea visto y, necesariamente, el sistema deberá "recordar" todos los resultados parciales hasta el último evento. En los gráficos de la **Figura 4** podemos ver la eficiencia de nuestra propuesta comparada con SASE y EsperTech, los dos referentes en la academia e industria, respectivamente. En el eje horizontal se muestra el número de datos procesados antes de ver el último evento, y en el eje vertical el uso de recursos en tiempo (gráficos de la izquierda) y memoria RAM (gráficos de la derecha). Como podemos ver, leyendo menos de dos mil datos (un número bastante bajo si pensamos en la cantidad de datos en un contexto real), los otros sistemas toman del orden de segundos para procesar Q_1 y del orden de minutos para procesar Q_2 . En cambio, nuestra propuesta, basada en algoritmos de enumeración con demora constante, toma tiempos del orden de milisegundos en ambas con-

sultas, muy por debajo del tiempo de los otros sistemas. Si vemos la cantidad de memoria usada durante los experimentos es incluso más dramático: los sistemas toman del orden de cientos a miles de megabytes de memoria, mientras que nuestra propuesta no alcanza los diez megabytes de memoria. De este experimento se pueden comprobar los beneficios de utilizar algoritmos de enumeración con demora constante. Los sistemas clásicos tratan de mantener todos los resultados parciales (que puede ser una cantidad exponencial) y tienen que actualizarlos todos cada vez que llega un nuevo evento. En cambio, nuestra representación compacta no necesita mucho espacio y, por otro lado, se puede actualizar rápidamente cada vez que llega un nuevo dato.

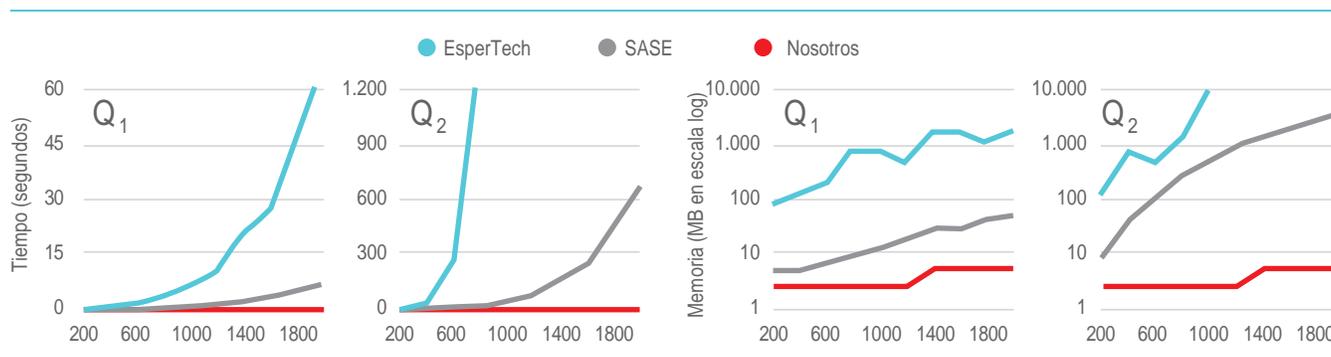


FIGURA 4 RESULTADOS DEL EXPERIMENTO DE STRESS EN NUESTRO SISTEMA, COMPARADO CON SISTEMAS DE LA INDUSTRIA (ESPERTECH) Y LA ACADEMIA (SASE).

CONCLUSIONES

COMO PUDIMOS OBSERVAR, EL MANEJO DE EVENTOS COMPLEJOS Y, MÁS GENERAL, EL MANEJO DE STREAMS DE DATOS, PROPONE GRANDES DESAFÍOS PARA LA CIENCIA DE LA COMPUTACIÓN, EN LA BÚSQUEDA DE ALGORITMOS EFICIENTES CAPACES DE PROCESAR GRANDES VOLÚMENES DE DATOS. NUESTRA PROPUESTA BUSCA SER UN PRIMER ACERCAMIENTO HACIA LA FORMALIZACIÓN DE SISTEMAS CEP, DEFINIENDO LAS BASES SOLIDAS PARA EL ESTUDIO DE ESTA ÁREA, Y DEJANDO VARIOS PROBLEMAS AÚN POR RESOLVER. ACTUALMENTE, NOS ENCONTRAMOS INVESTIGANDO NUEVOS ALGORITMOS DE ENUMERACIÓN CON DEMORA CONSTANTE PARA INCLUIR OTRAS CARACTERÍSTICAS DE ESTOS SISTEMAS, COMO MANEJO DE VENTANAS DE TIEMPO, CORRELACIÓN ENTRE LOS DATOS, Y CÁLCULO DE ESTADÍSTICAS SOBRE LOS DATOS. POR ÚLTIMO, TAMBIÉN NOS ENCONTRAMOS IMPLEMENTANDO, JUNTO A UN GRUPO DE ALUMNOS, UN PRIMER PROTOTIPO DE NUESTRA PROPUESTA, EL CUAL ESPERAMOS QUE ESTÉ LISTO EN EL CORTO PLAZO PARA SER PROBADO EN ESCENARIOS REALES. ■

LA DEMOCRACIA EN LA ENCRUCIJADA DIGITAL





MARTIN HILBERT

Profesor de la Universidad de California, Davis, Estados Unidos. Doctor en Comunicación, Universidad del Sur de California, Estados Unidos, y Doctor en Economía y Ciencias Sociales, Universidad de Erlangen-Núremberg, Alemania.

Su investigación aborda un enfoque multidisciplinario que busca comprender el rol de la información y el conocimiento en el desarrollo de sistemas socio-tecnológicos complejos. Creador y coordinador del Programa Sociedad de la Información de CEPAL. En sus 15 años como Oficial de Asuntos Económicos de las Naciones Unidas, brindó asistencia técnica en el área de desarrollo digital a presidentes, expertos gubernamentales, legisladores, diplomáticos, ONGs y empresas, en más de 20 países.

hilbert@ucdavis.edu
www.martinhilbert.net

La revolución digital actualmente está destruyendo la democracia como la conocemos. Como resultado, la democracia está bajo amenaza en todo el mundo. Al mismo tiempo, las tecnologías digitales ofrecen el potencial de construir una “democracia más democrática” que cualquiera que hayamos visto en la historia. Necesitamos actuar rápidamente para convertir la amenaza actual en una oportunidad.

LA TECNOLOGÍA SIEMPRE HA ROTO Y HABILITADO LA DEMOCRACIA

La vigilancia, la manipulación de opiniones, los movimientos ocultos de preferencias, los ejércitos de robots de opinión, las burbujas de filtro y las cámaras de eco están en desacuerdo con los principios fundamentales de la democracia representativa. Se ha demostrado reiteradamente que han influido en elecciones y opiniones políticas [Boichak, Jackson, Hemsley, & Tanupabrungsun, 2018; Hilbert, Ahmed, Cho, Liu, & Luu, 2018; Kupferschmidt, 2017], y han intensificado la polarización en un grado que parece difícil de reconciliar con la ambición de crear una perspectiva común [Bakshy, Messing, & Adamic, 2015; Pariser, 2011].

La democracia es, en esencia, un problema de procesamiento de la información. Ésta es la razón por la cual estas tecnologías de procesamiento de información afectan el proceso democrático con tanta profundidad. Como resultado, en los últimos años se ha visto un cambio hacia los líderes autocráticos, muchos de ellos elegidos democráticamente, pero que están trabajando en el desmantelamiento de las tradiciones democráticas, a menudo con el uso explícito de herramientas digitales. El principal culpable no es la tecnología digital *per se*, sino el hecho que no hemos construido la tecnología adecuada y que no hemos ajustado nuestras instituciones democráticas.

La implementación de la democracia siempre ha ido de la mano con las posibilidades tecnológicas disponibles. Se le atribuye a Aristóteles el decir que la democracia solo se extiende

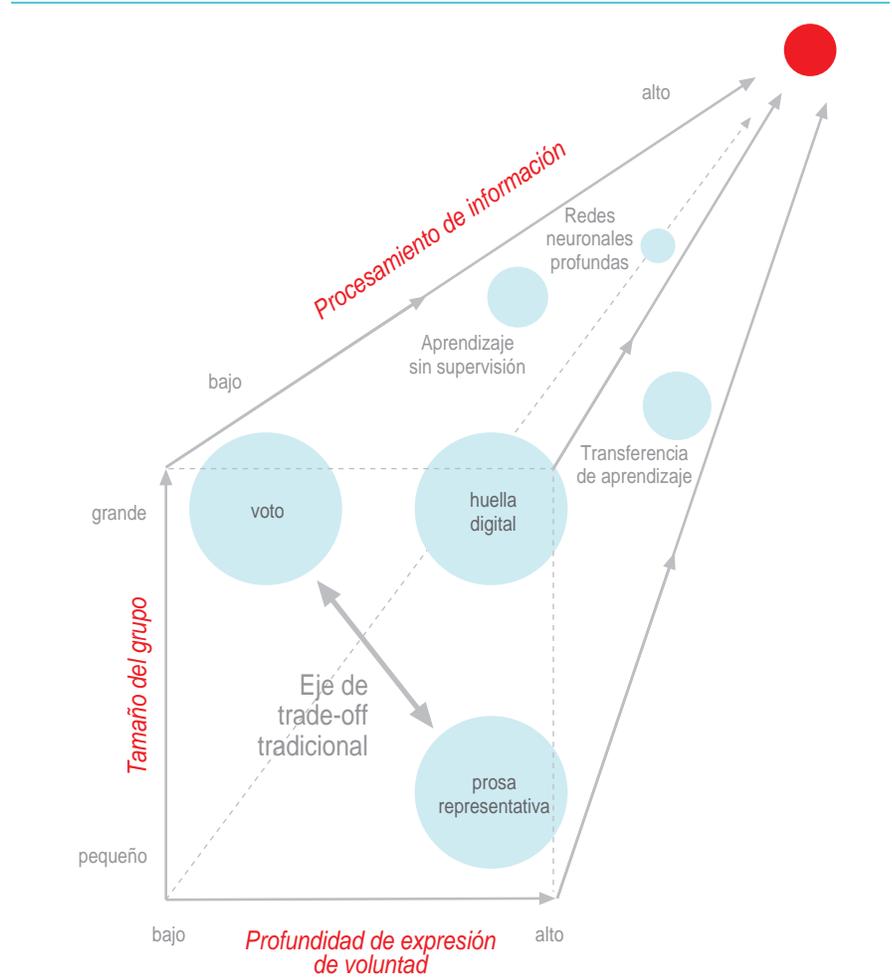


FIGURA 1. REPRESENTACIÓN ESQUEMÁTICA DEL TRADE-OFF ENTRE EL TAMAÑO DEL GRUPO INVOLUCRADO EN LAS ETAPAS DE DELIBERACIONES Y TOMA DE DECISIONES, Y LA PROFUNDIDAD DE SU EXPRESIÓN DE VOLUNTAD.

hasta donde un hombre puede caminar en un día, ya que la información no podía viajar más distancia en un tiempo razonable. La democracia de Atenas se limitaba a una ciudad. Cuando Abraham Lincoln fue elegido en 1860, los ciudadanos de California tardaron cinco días en saber quién era el presidente, ya que el Pony Express tenía que llevar el mensaje por todo el país. Si bien la información de hoy viaja a la velocidad de la luz, las instituciones democráticas actuales siguen siendo muy similares a lo que han sido durante siglos. No debería sorprender que una revolución de la tecnología de la información y la comunicación, tan profunda como la revolución digital, haga obsoleta las instituciones democráticas, construidas para

una época donde la información, viajaba a la velocidad de un caballo. No debería sorprendernos que segmentos específicos de la sociedad aprendan a aprovechar el desajuste actual entre nuestra realidad digital y nuestras instituciones democráticas, para impulsar su propia agenda antidemocrática.

EL COMPROMISO DE TAMAÑO-PROFUNDIDAD

Históricamente, el desafío de la democracia ha sido tratar con demasiada información. Nuestro sistema actual de votación y democracia representativa es la respuesta al desafío

histórico de mediar entre la escala, o tamaño del grupo, y la profundidad de la expresión de la voluntad. El procesamiento digital de la información de hoy mantiene la promesa de romper este compromiso o *trade-off*. Si bien la idea de procesar las preferencias cualitativamente expresables de 7.500 millones de personas en todo el mundo parece desalentadora, la inteligencia artificial, como las *deep neural networks* o redes neuronales profundas, solo es efectiva cuando se la capacita en cientos de miles de millones de casos. Con la información digital, el big data no es una limitación, sino una oportunidad. El desafío consiste en crear tecnologías que puedan reconciliar el equilibrio entre la profundidad y la escala.

La **Figura 1** proporciona una esquematización de las ideas principales detrás del desafío tradicional de la democracia y el camino potencial hacia adelante con los medios digitales [basado en Hilbert, 2009]. En la práctica, las restricciones de procesamiento de la información no han permitido la participación significativa de todos. En respuesta, una gran variedad de diversos mecanismos y procedimientos han evolucionado a lo largo de los siglos. La mayoría de ellos básicamente giran en torno al conocido eje entre la democracia directa y la democracia representativa. La democracia representativa reduce el número de ciudadanos directamente involucrados en varias etapas de la deliberación. Un pequeño grupo, a menudo denominado "filtro de Madison" [Madison, 1787], está legitimado por el resto y lleva a cabo un discurso profundo en representación de la gente. La alternativa, la democracia directa, no reduce el número de participantes, sino que simplifica la expresión de opiniones. En lugar de entrar en un discurso de influencia mutua, un gran número de personas vota sobre un conjunto restringido de temas predefinidos. En la democracia moderna, el canal resultante comunica casi exactamente 1 bit de información del ciudadano en búsqueda de su voluntad democrática: izquierda o derecha. Éste es un ancho de banda vergonzosamente pequeño a la luz de las posibilidades de hoy.

La **Figura 1** ilustra la lógica detrás de la antagoría entre la opción de un pequeño cuerpo de

	Tamaño del grupo	Profundidad de expresión de voluntad	Procesamiento de información
Voto sí / no	↑	↓	↓
Prosa representativa	↓	↑	↓
Huella digital	↑	↑	↓
Democracia digital	↑	↑	↑

FIGURA 2
CLASIFICACIÓN DE LOS DIVERSOS MECANISMOS DE DEMOCRACIA DE ACUERDO AL TAMAÑO DEL GRUPO, PROFUNDIDAD DE EXPRESIÓN DE VOLUNTAD Y PROCESAMIENTO DE INFORMACIÓN QUE PERMITEN.

ciudadanos que filtran la opinión pública a través de discursos complicados y ambiguos (prosa representativa), y la alternativa de expresiones breves y predefinidas, pero indiscutiblemente expresiones de la mayoría (voto directo sí/no). El desafío actual de la e-democracy agrega una tercera dimensión a esta lógica originalmente bidimensional, que consiste en el procesamiento automático de la información. Las tecnologías que tienen tal vez el mayor potencial son las llamadas redes neuronales profundas [Goodfellow, Bengio, & Courville, 2016; LeCun, Bengio, & Hinton, 2015]. En la representación democrática, el objetivo es convertir las opiniones no estructuradas de las masas en resultados viables, canalizándolas a través de diferentes capas de representación (ya sea geográfica, cultural, de opinión o de perspectiva). El llamado aprendizaje profundo es sorprendentemente bueno para hacer exactamente eso, es decir, para equilibrar la tensión entre preservar la mayor cantidad de información posible sobre los datos originales y, al mismo tiempo resultar útil para la transformación de la información, incluso, no estructurada. En las redes neuronales profundas, esto se implementa a través de una red de múltiples capas (similar a las instituciones democráticas), en la que las unidades que no están en la capa de entrada o salida se denominan convencionalmente *unidades oculatas* (irónicamente, eso también es similar a las

instituciones de la democracia representativa...).

Las tecnologías modernas tienen como objetivo cerrar la brecha que surge por el antagonismo entre el tamaño del grupo y la profundidad de su expresión de voluntad. Como se muestra en la **Figura 2**, el objetivo es combinar la fortaleza de los grandes datos proporcionados por la huella digital, con decisiones procesables. Algunas soluciones apuntan a facilitar la "camisa de fuerza" de simplificación en torno a las opciones ofrecidas para votar, mientras que otras apuntan a permitir que más y más ciudadanos participen en la maraña impenetrable de las deliberaciones democráticas. El primer grupo apunta a aumentar el número de opciones cuantitativas, mientras que el segundo apunta a extraer ideas estructuradas de expresiones cualitativas.

IDEAS ACCIONABLES DESDE MÁS OPCIONES

Aumentar el número de opciones es tan fácil como proporcionar un cuestionario más complejo que la opción típica "izquierda/derecha" ofrecida en las elecciones de hoy. Las llamadas "apps de asesoramiento de votación" son una implementación de esto: los ciudadanos res-

ponden una serie de 20-50 preguntas simples, que luego se comparan con los programas de los partidos políticos para sugerir cómo votar. Los votos binarios de hoy siguen siendo un recurrente claro desde los tiempos en que no era factible ir más profundo computacionalmente, y si no llegan a un resultado claro, se establecen otros mecanismos institucionales, como los votos de segunda vuelta y las segundas rondas. Sin embargo, es bien conocido, a partir de los trabajos de los matemáticos Condorcet y Borda del siglo XVIII, que existen formas más elegantes y efectivas de determinar ganadores democráticos, simplemente mediante la explotación de más información. Borda, por ejemplo, ideó una escala que clasificaba las preferencias según la fuerza con la que se tenían. Curiosamente, al obtener preferencias de rango detalladas de los ciudadanos, el ganador casi nunca es una de las primeras opciones de la mayoría de todos los ciudadanos, sino un candidato de consenso, más abajo en el rango de todos. En este sentido, las instituciones existentes, como la segunda ronda de votación, prefieren enfrentar entre sí a los grupos de ciudadanos recompensando a la facción con un 50,1% y castigando a la otra facción con un 49,9%, en lugar de identificar un tercer candidato de consenso con el que todos los ciudadanos estarían conformes.

La combinación de una papeleta de votación informativa más rica con capacidades de aprendizaje automático conduce a otros beneficios potenciales. Podemos obtener información nueva altamente no intuitiva de las llamadas técnicas de aprendizaje no supervisadas. El análisis de componentes principales e independientes (PCA e ICA, respectivamente) se encuentra entre los algoritmos de aprendizaje de representación más antiguos, que se remonta a más de un siglo [Pearson, 1901]. Estos y otros algoritmos de aprendizaje relacionados (como el análisis factorial) esencialmente aprenden una variedad con regiones delgadas de alta probabilidad a lo largo de algunos ejes, y alargamientos a lo largo de otros, como una forma de panqueque [Hinton, Dayan, & Revow, 1997]. Uno de los logros iniciales de las redes neuronales profundas fue encontrar generalizaciones no lineales de PCA [Hinton & Sa-

lakhtudinov, 2006]. Estos algoritmos de aprendizaje también se pueden usar para desentrañar las opiniones y las estructuras de preferencias de diferentes grupos de ciudadanos, e incluso llevar a nuevas perspectivas sobre la estructura de opinión existente.

Por ejemplo, creamos un cuestionario que preguntaba a 126 estudiantes acerca de las preferencias sobre ocho temas relacionados con "Obamacare" y la contrapropuesta, llamada "Trumpcare". Aplicamos un algoritmo de aprendizaje no supervisado basado en PCA para analizar la distribución de esta voluntad popular. Una simulación de Monte Carlo de valor propio (también conocida como "análisis paralelo") sugirió identificar cuatro componentes dentro de nuestras ocho preferencias de políticas de rango ponderado. En conjunto, los cuatro componentes explicaron el 63% de la varianza total. Resultó que cada uno de los cuatro componentes estaba dominado principalmente por dos cargas, y que esto distribuía los ocho números de manera excluyente y exhaustiva. Además, y de manera bastante interesante, resultó que las dos cargas dominantes de cada componente siempre consistían en opciones de políticas opuestas: si una persona prefería la opción de Obamacare en un tema, lo que maximizaba la varianza de este componente era relativamente más apoyo para la opción Trumpcare en otro problema dominante (o viceversa). En otras palabras, en lugar de identificar los problemas que se encuentran en concordancia con los partidos políticos, los componentes principales cruzaron las líneas partidarias tradicionales de manera consistente, identificando los componentes que contrastaban las preferencias de cada parte.

Sobre la base de este resultado, creamos cuatro opiniones políticas estilizadas, que encarnan las estructuras de preferencias identificadas entre los partidos. Lo compartimos con los participantes, con el siguiente mensaje: "Después de la primera ronda de encuestas, las primarias han decidido que cuatro representantes políticos continuarán trabajando en la futura reforma de la atención médica. Los cuatro candidatos tienen las siguientes prioridades (informado por nuestra PCA). Por favor, clasi-

fíquelos desde el representante que mejor lo represente en estas deliberaciones, al que menos represente su opinión". El 89% de los participantes de nuestra primera ronda participaron en esta segunda ronda dos semanas después. Los participantes identificaron a dos de los cuatro representantes como los más representativos de sus preocupaciones. La opción que obtuvo el puntaje más alto optó por Obamacare cuando se trata del tema de la Salud de la Mujer, y por Trumpcare cuando se trata de MEDICAID (contra el apoyo a grupos de bajos ingresos). También les pedimos nuevamente que repitieran la encuesta original, y nuevamente detectamos este rasgo específico en la estructura de opinión del grupo. Esto reafirma que encontramos al menos una estructura de preferencias oculta importante en nuestro pequeño grupo de ciudadanos. Aquellos que se sienten fuertemente a favor de la opción de Obamacare para la Salud de la Mujer, se sienten menos favorables por la opción MEDICAID de Obamacare y favorecen la opción de Trumpcare, y viceversa. Esto no es de ninguna manera intuitivo (al menos no lo fue para nosotros), pero dentro de nuestro grupo ha sido revelado de manera consistente por el algoritmo de aprendizaje no supervisado.

Este hallazgo proporciona información importante para las negociaciones democráticas en torno al debate político en curso. Los votantes republicanos se molestaron cuando la tiranía de la mayoría favoreció la opción Obamacare de Salud de la Mujer, lo que lleva a resultados democráticos inestables. Varios de los temas involucrados han sido muy polémicos (incluido el financiamiento para abortos). Gracias al aprendizaje no supervisado, ahora sabemos que este acuerdo podría fortalecerse ofreciendo trabajar con la propuesta republicana en torno a MEDICAID. En promedio, hemos visto que aquellos que tienen como prioridad a Obamacare para la Salud de la Mujer, no tienen un fuerte apoyo de Obamacare para MEDICAID. Esto permite dar algo a la mayoría, y al mismo tiempo, ofrecer algo a la oposición.

Por supuesto, la intermediación de este acuerdo no satisficará a todos. Son estadísticas construidas a partir de promedios. También es impor-

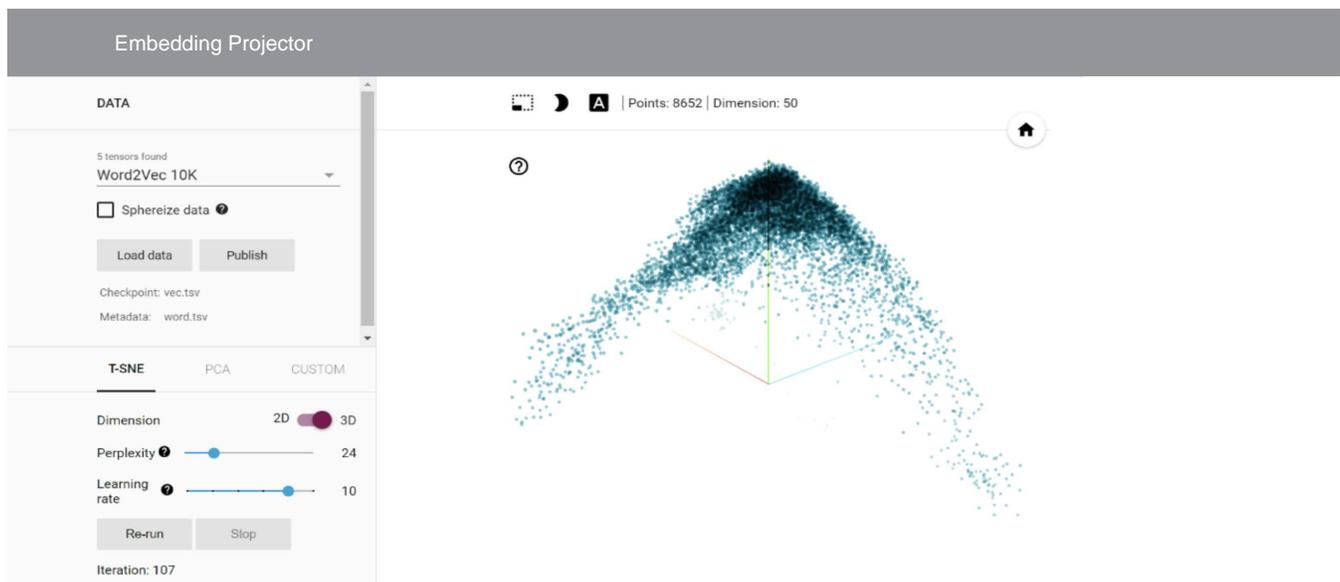


FIGURA 3
ESPACIO VECTORIAL WORD2VEC EXTRAÍDO A PARTIR DE 3.000 ARTÍCULOS PERIODÍSTICOS SOBRE LA REFORMA DEL SISTEMA DE SALUD ESTADOUNIDENSE.

tante enfatizar que no siempre es posible encontrar un óptimo que satisfaga a todos. Éste es un hecho matemático, debido a la multidimensionalidad de las estructuras de preferencias [Arrow, 1950]. Sin embargo a menudo se pueden encontrar óptimos locales, que satisfacen a la mayoría de las personas, y esto es democráticamente deseable. La nueva percepción podría ayudar a redactar proyectos de ley que disminuyan el número de ciudadanos frustrados y creen un resultado democrático más sostenible.

MAPEO DE ESPACIOS DE OPINIÓN MULTIDIMENSIONALES

Yendo un paso más allá, no solo podemos obtener los datos en bruto de la expansión de la papeleta de votación de una elección binaria a una encuesta larga, como se hizo en el caso anterior, sino el objetivo de extraer ideas procesables de expresiones de opinión puramente cualitativas. En lugar de pedirles a las personas que elijan entre opciones preformuladas, que a menudo ya distorsionan la voluntad de las personas, uno le pedirá a cada ciudadano que

simplemente exprese su opinión, de la manera que ellos deseen. Esto sería luego alimentado a una red neuronal profunda, que nos proveerá el espacio de opinión multidimensional que representa la realidad multidimensional de las opiniones democráticas mucho mejor que una perspectiva unidimensional de izquierda/derecha como fue realizado, por ejemplo, durante la consulta constitucional en Chile en 2016 [Fierro, Fuentes, Pérez, & Quezada, 2017].

La **Figura 3** visualiza este desafío en forma de un espacio vectorial Word2Vec [Mikolov, Sutskever, Chen, Corrado, & Dean, 2013], que creamos con la herramienta de acceso abierto de Google <http://projector.tensorflow.org/>. Recopilamos unos 3.000 artículos de opinión periodísticos sobre la reforma del sistema de salud estadounidense de Washington Post, CNN y Fox News, y creamos el siguiente espacio vectorial en 50 dimensiones (usando T-SNE para representarlo en 3D). Podemos ver claramente una forma de este espacio, y esperamos que éste nos permita comprender la estructura de opinión subyacente.

Este espacio fue creado mediante la conversión de palabras a vectores. Como ahora tratamos

con vectores, también podemos realizar las operaciones habituales de álgebra lineal en el espacio vectorial. "Por ejemplo, la relación hombre / mujer se aprende automáticamente, y con las representaciones vectoriales inducidas, "Rey - Hombre + Mujer" da como resultado un vector muy cercano a "Reina" [Mikolov, Yih, & Zweig, 2013]. En el mismo sentido, uno podría evaluar cuáles de los vectores que presentan intereses de un partido político se alinean más fuertemente con los intereses del partido opositor, y cuáles menos. Esto proporciona información importante para buscar soluciones de consenso al redactar proyectos de ley o al redactar mensajes políticos o agendas.

Es importante darse cuenta de que cada espacio vectorial creado es el resultado del conjunto de datos de entrenamiento subyacente. Experimentamos con algunas opciones y encontramos diferencias sensibles. Por ejemplo, si a nuestro espacio vectorial de la atención médica basado en 3.000 opiniones le damos cuatro términos ("hombre", "mujer", "inhumano", "Kevin") y calculamos cuál es el menos alineado con los demás, identificamos a "Kevin" como el menos alineado. Luego, entrenamos un espacio vectorial basado en un conjunto de

datos genérico de 3GB de Google y le hicimos la misma pregunta. Esta vez, identifiqué "inhumano" como el menos alineado. Ambas respuestas tienen sentido. "Kevin" es el único nombre, mientras que "inhumano" es el único término que describe algo que no es humano. Si nuestra ambición es utilizar dichas herramientas para el discurso democrático, debemos ser conscientes de estas diferencias sutiles. Queda, aún, mucho trabajo por hacer en este campo.

DE LOS POLÍTICOS A LOS ALGORITMOS

En resumen, el progreso tecnológico en el ámbito de la información y la comunicación llevó a un desajuste entre nuestra realidad tecnológica y nuestros procesos democráticos. Esto no es culpa de la tecnología en sí, sino de la tecnología que empleamos actualmente. La gran mayoría de las inversiones en el desarrollo del aprendizaje automático moderno y la inteligencia artificial tienen dos propósitos:

optimizar el marketing (por ejemplo, en Silicon Valley) y optimizar la seguridad nacional (por ejemplo, por la NSA, CIA, etc.). La adopción de las tecnologías resultantes con fines democráticos conduce a la creciente destrucción de la democracia. ¿Por qué debe una tecnología diseñada para optimizar la correspondencia económica entre la oferta y la demanda fomentar, también, el discurso democrático? ¿Por qué debe una tecnología diseñada para mantener a los ciudadanos seguros fomentar, también, las libertades democráticas individuales? Sería sorprendentemente agradable si fuera el caso, pero por el momento es algo demasiado ambicioso, y adoptar estas tecnologías para fines democráticos no ha funcionado.

Hasta ahora, no hay inversiones considerables para crear tecnología digital que realmente fomente la democracia. Sin embargo, en principio, dicha tecnología puede ser creada. Al mismo tiempo, será necesario reformar nuestras instituciones democráticas. Esto significará volver a escribir nuestras Constituciones. En una versión extrema de la visión propuesta,

una red neuronal profunda podría incluso reemplazar gran parte de lo que está haciendo el Congreso. En palabras de Madison: tomar la opinión pública en su "forma bruta" y "refinar y ampliar las opiniones del público, pasándolas a través de un medio" [Madison, 1787]. El medio no sería un grupo de personas, sino una red neuronal.

Por supuesto, uno no debería imaginarse el resultado final como una máquina mágica alimentada con la voluntad del pueblo, que luego produce leyes sólidas. Pero de la misma manera que las máquinas reemplazan actualmente gran parte del trabajo de abogados, periodistas, y analistas financieros, la inteligencia artificial también ofrece la posibilidad de reemplazar gran parte del trabajo realizado por representantes democráticos. La principal tarea de los representantes democráticos es sintetizar, expandir, refinar y ampliar la voluntad de los ciudadanos mediante el procesamiento de la información. Las máquinas son eventualmente mejores en eso. Y esto es algo bueno. Desde las revoluciones francesa y estadouni-



dense en el siglo XVII, ha sido la ambición de las democracias modernas reemplazar el "gobierno del hombre", que está sujeto a sesgos arbitrarios y de interés propio, con el objetivo del "estado de derecho" y el "imperio de la ley". Por definición, los algoritmos son el epítome del "imperio de la ley". Los algoritmos son la

encarnación del "imperio de la ley". La estructura de preferencia de los algoritmos no está tácitamente oculta o sesgada por intereses personales. Puede (¡y debe!) abrirse y verificarse de una manera muy transparente, mientras que, al mismo tiempo, procesa mucha más información de la que cualquier cerebro biológico podría

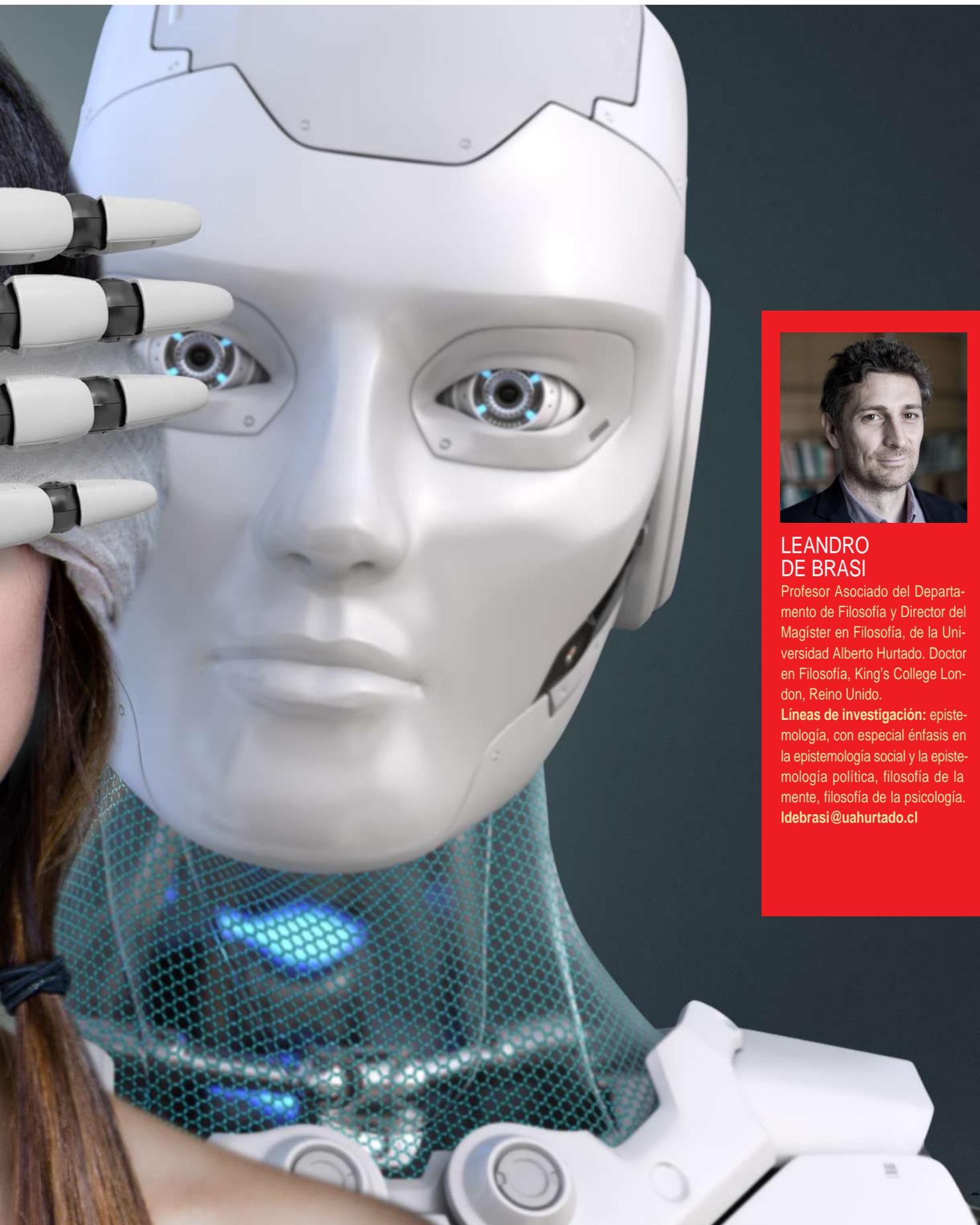
manejar. Como tales, nuestras democracias modernas solo pueden beneficiarse de la transferencia a algoritmos inteligentes de algunas tareas que hoy en día ejecutan nuestros políticos. ■

REFERENCIAS

- Arrow, K. J. (1950). A Difficulty in the Concept of Social Welfare. *Journal of Political Economy*, 58(4), 328–346.
- Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239), 1130–1132.
- Boichak, O., Jackson, S., Hemsley, J., & Tanupabrungsun, S. (2018). Automated Diffusion? Bots and Their Influence During the 2016 U.S. Presidential Election. In G. Chowdhury, J. McLeod, V. Gillet, & P. Willett (Eds.), *Transforming Digital Worlds* (pp. 17–26).
- Fierro, C., Fuentes, C., Pérez, J., & Quezada, M. (2017). 200K+ Crowdsourced Political Arguments for a New Chilean Constitution. In *Proceedings of the 4th Workshop on Argument Mining* (pp. 1–10). Copenhagen, Denmark: Association for Computational Linguistics.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Hilbert, M. (2009). The Maturing Concept of E-Democracy: From E-Voting and Online Consultations to Democratic Value Out of Jumbled Online Chatter. *Journal of Information Technology & Politics*, 6(2), 87.
- Hilbert, M., Ahmed, S., Cho, J., Liu, B., & Luu, J. (2018). Communicating with Algorithms: A Transfer Entropy Analysis of Emotions-based Escapes from Online Echo Chambers. *Communication Methods and Measures*, 0(0), 1–16.
- Hinton, G., Dayan, P., & Revow, M. (1997). Modeling the manifolds of images of handwritten digits. *IEEE Transactions on Neural Networks*, 8(1), 65–74.
- Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the Dimensionality of Data with Neural Networks. *Science*, 313(5786), 504–507.
- Kupferschmidt, K. (2017). Bot-hunters eye mischief in German election. *Science*, 357(6356), 1081–1082.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
- Madison, J. (1787). The Federalist No. 10 The Utility of the Union as a Safeguard Against Domestic Faction and Insurrection (continued). *Daily Advertiser*, Thursday, November 22. Retrieved from <http://www.constitution.org/fed/federa10.htm>
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. In *Advances in Neural Information Processing Systems* (pp. 3111–3119).
- Mikolov, T., Yih, S. W., & Zweig, G. (2013). Linguistic Regularities in Continuous Space Word Representations. Microsoft Research.
- Pariser, E. (2011). *The Filter Bubble: What the Internet Is Hiding from You*. Penguin.
- Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine Series 6*, 2(11), 559–572.



PERSONALIZACIÓN,
POSTVERDAD Y
DEMOCRACIA:
EFECTOS DE LAS
TECNOLOGÍAS
DIGITALES DE LA
INFORMACIÓN EN
LA CIUDADANÍA



LEANDRO DE BRASI

Profesor Asociado del Departamento de Filosofía y Director del Magíster en Filosofía, de la Universidad Alberto Hurtado. Doctor en Filosofía, King's College London, Reino Unido.

Líneas de investigación: epistemología, con especial énfasis en la epistemología social y la epistemología política, filosofía de la mente, filosofía de la psicología. idebrasi@uahurtado.cl

La idea que la democracia, como “gobierno del pueblo”, es la única forma de gobierno legítima es reciente y pasó a ser ampliamente aceptada solo desde la segunda mitad del siglo XX. Pese a esto, durante ese período distintas democracias en distintos países, incluyendo Chile, sufrieron en manos de aquellos que portan armas. Pero las democracias no solo pueden morir abruptamente en golpes de estado. También pueden morir más lentamente por medio de gobiernos electos por una ciudadanía desinformada, polarizada y segregada.

Elecciones, aunque sean justas y libres, no son suficientes para la democracia. Ésta también requiere que sus ciudadanos sean mínimamente competentes. Para eso, los ciudadanos necesitan por lo menos estar bien informados sobre distintos asuntos (por ejemplo, económicos, educacionales, migratorios, etc.) y lo suficientemente capacitados para asimilar y reflexionar sobre esa información y así ser capaces de formar opiniones razonables. Por supuesto que existe una larga tradición antidemocrática, que surge casi inmediatamente con la formación de las primeras democracias en la antigua Grecia y que niega que la ciudadanía satisfaga esta competencia mínima. Por ejemplo, en el *Gorgias*, Platón afirma que la democracia es defectuosa dada la ignorancia de sus ciudadanos y, dado esto, sugiere, en la *República*, una epistocracia: el gobierno de los que más saben. Este tipo de preocupación acerca de la incompetencia ciudadana es planteada por todo tipo de teóricos políticos (véase, por ejemplo, [1, 2, 3]). Gran parte de esta preocupación se basa en la desinformación y segregación epistémica de los ciudadanos y la polarización de sus opiniones.

Ahora bien, uno quizá piense que las nuevas tecnologías digitales de la información, que en sus comienzos prometieron más información, expresión, comunicación y conexión, entre otras cosas, podrían beneficiar a la democracia: en particular, podrían ayudar a superar estas preocupaciones. Pero en reali-

dad, como veremos a continuación, ciertas tecnologías digitales magnifican el problema (uno podría haber esperado esto ya que las tecnologías son normalmente puestas en uso antes de ser completamente entendidas y muy raramente somos capaces de predecir sus consecuencias en sus comienzos). De hecho, estas tecnologías, como otras en el pasado (por ejemplo, la prensa y la televisión), introducen un nuevo desafío epistémico a la competencia ciudadana. Dado que estas tecnologías de la información son (y probablemente serán aún más) extensamente utilizadas, especialmente por “nativos digitales”, debemos aprender a interactuar con ellas si no queremos que las futuras democracias sean solo democracias en nombre.

ALGORITMOS DE PERSONALIZACIÓN Y DEFICIENCIAS COGNITIVAS

Asumo que no necesito convencer al presente lector acerca de la gran mediación digital a la cual muchas de nuestras vidas están sujetas. Aquí me concentraré en algunos aspectos de algunas tecnologías digitales de la información. En particular, me concentraré en buscadores, como Google, y redes sociales como Facebook, que son ampliamente utilizados, para ilustrar las consecuencias epistémicas negativas que estas tecnologías tienen en consumidores de noticias online. No asumo que las compañías detrás de estas tecnologías tienen malas intenciones: por ejemplo, intenciones de corromper la infoesfera. Por supuesto que a toda tecnología (como muchas otras cosas) se le puede dar un mal uso (compañías como Cambridge Analytica parecen estar haciéndolo [4]), pero ése no es un asunto que me interesa considerar en esta oportunidad. En cambio, me interesa contemplar cómo nuestra interacción con estas tecnologías está cambiando radicalmente la manera en que nos informamos y formamos nuestras creencias.

Hay cada vez más información disponible en la Web. Pero la disponibilidad de tanta infor-

mación genera una demanda de atención que no podemos, en términos prácticos, satisfacer. Por eso, Google y otros buscadores tienen la función de dirigir nuestra atención hacia aquella parte de la Web que es más relevante a nuestros intereses. Para esto, los buscadores deben filtrar la información de alguna manera. Y la manera por la cual está siendo filtrada, desde el año 2010, por prácticamente todo servicio de Internet (incluyendo servicios de venta y entretenimiento), es de acuerdo a las preferencias e intereses de cada usuario en particular. Lo que empuja esta personalización del servicio es una visión del mundo digital hecho a la medida de cada uno. Esta personalización es importante para estas compañías porque los avisos que uno recibe, al utilizar sus servicios, también están dirigidos de manera personalizada (recuerde que si uno no paga por un buen servicio online, lo más probable es que uno sea el producto que se vende). Dada la información que estas compañías recogen cada vez que sus servicios son utilizados por sus usuarios, pueden hacer que los avisos sean más relevantes para uno y así vender a terceros más espacios publicitarios dirigidos especialmente a cada usuario según sus preferencias e intereses.

Pero esta información también es explotada por algoritmos de personalización que filtran los contenidos de la Web según las preferencias e intereses particulares de cada uno. Ahora bien, dejando de lado temas de privacidad (dada la información que estos servicios deben recolectar acerca de uno para poder personalizarlos), el filtrado por personalización no es algo malo en sí mismo. Cuando busco entretenerme con alguna película de Netflix o video de YouTube, o comprar algún libro en Amazon, estos servicios me sugieren cosas que podrían ser de mi agrado dada la información que poseen acerca de mí. El problema surge cuando uno busca informarse por medios digitales que explotan esta personalización. En este caso, el filtrado personalizado llevará mi atención a noticias que son de mi preferencia, y por supuesto que tales noticias no necesariamente reflejan la realidad.

De esta manera, las dietas de información de mucha gente (por ejemplo, un reciente estudio del Reuters Institute de la Universidad de Oxford encontró que, en Chile, 89% de los usuarios online utiliza el medio digital, incluyendo las redes sociales, para informarse [5]) se podría ver seriamente afectada por estos algoritmos, dado que no toda información disponible en los medios digitales es correcta o por lo menos producida responsablemente. De los 2.5 trillones de bytes de data diarios que son actualmente creados, preocupantemente mucha información es intencionalmente engañosa o simplemente falsa, inclusive que se ajusta a mis preferencias e intereses [6]. Y aunque siempre ha habido gente que quiere engañarnos, el medio digital con sus algoritmos de personalización hace que el trabajo de separar la buena de la mala información, y así entonces de formar creencias razonables acerca de un sinfín de asuntos, sea más complicado aún. Pero note que el problema no es solo que la dieta de información de uno podría estar contaminada. Lo realmente preocupante no es eso, sino que, dada esta dieta restringida por mis preferencias, muy probablemente no tenga acceso a información que contrarreste la información engañosa o falsa: en otras palabras, acceso a buena información que me permita cuestionar la mala información que se acomoda a mis preferencias. Así entonces vamos creando nuestras propias “burbujas digitales”, donde las cosas parecen ser como queremos que sean y no necesariamente como son.

Estos algoritmos hacen que el trabajo epistémico de separar la buena de la mala información sea aún más laborioso ya que exacerban ciertas deficiencias cognitivas. Por ejemplo, considere el sesgo de confirmación, reconocido hace ya mucho tiempo y que todos poseemos [7]. Ésta es la tendencia natural que tenemos a buscar y encontrar información que apoya a nuestras creencias e ignorar información que las contradice. El algoritmo de personalización justamente potencia este sesgo cognitivo, ya que nos presenta solo información que se acomoda a nuestras creencias. Además,

la gente siente placer cuando procesa información que sustenta sus creencias [8]. Este componente fisiológico que posee el sesgo de confirmación lo hace particularmente peligroso y los medios digitales que personalizan la información parecen estar diseñados para proveer un torrente de dopamina cuando ojeamos las noticias. Por lo tanto, estos algoritmos fomentan la colección sesgada de información y hacen que sea menos probable que dejemos de creer lo que ya creemos equivocadamente.

Dado lo anterior, la personalización también tiene el potencial de polarizar las creencias y hacer que la infoesfera sea más amigable a las “noticias falsas” (noticias intencionalmente falsas). De hecho, tanto las redes sociales como los buscadores que explotan la personalización organizan nuestra exposición a la información de tal manera que no nos permiten tener acceso a posiciones contrarias sofisticadas. En cambio, es probable que uno solo tenga acceso (a través de fuentes que simpatizan con la posición de uno) a caricaturas de esas posiciones (que son fácilmente refutadas). Eso entonces hace que cada vez más uno entre en conversaciones con gente como uno, cuyas opiniones reafirman las creencias de uno, exacerbando así nuestro sesgo de sobreconfianza [9]. Este sesgo consiste en la sobreestimación de la certeza de las creencias propias (en otras palabras, tenemos una tendencia a pensar que nuestras creencias son más seguras, epistémicamente hablando, de lo que realmente lo son). En este caso, el algoritmo de personalización exagera este sesgo cognitivo dado que no nos expone a posiciones sofisticadas distintas a las nuestras, creando así “cámaras de eco”, donde las mismas opiniones se repiten, y provoca que uno tenga más confianza en sus creencias de la que debería tener. Todo esto hace que uno demasiado fácilmente se convenza de que lo que cree es realmente conocimiento, o por lo menos creencias muy bien justificadas, cuando de hecho no lo son. De esta manera, entonces, estos algoritmos diseñados para servirle a nuestras preferencias, amplifican la polarización de opiniones y proveen un entorno amigable a la desinformación.

Por lo tanto, existe un claro riesgo de vivir en burbujas digitales que ratifican nuestras creencias, correctas o no, y en cámaras de eco que nos hacen sentir como que ya sabemos todo, aunque estemos críticamente desinformados. Pero el Internet con su tecnología de personalización no es solo un gran mecanismo de consolidación de creencias que ya estamos sesgados a creer, sino que también alienta activamente una cierta arrogancia epistémica que promueve la intolerancia. De hecho, una reciente encuesta, realizada a jóvenes estudiantes de universidades norteamericanas, arrojó resultados alarmantes en relación a los niveles de intolerancia que estos poseen [10]. El 51% de los jóvenes estaba de acuerdo con perturbar el desarrollo de una charla a gritos para que la audiencia no escuche su contenido si no estaban de acuerdo con las ideas del orador. Más alarmante aún, el 19% estaba de acuerdo con utilizar violencia para prevenir que el orador exponga sus ideas. La mezcla tóxica de algoritmos personalizados y sesgos cognitivos tiene la capacidad de generar una segregación epistémica en la cual distintos grupos no solo se separan y no dialogan, también creen que los otros no entienden y por lo tanto sus posiciones no deben ser escuchadas.

VIRTUDES INTELLECTUALES Y POSTVERDAD

Para combatir esta situación, que no nos ayuda a diferenciar lo verdadero de lo falso, debemos ser epistémicamente despabilados. Y para poder serlo, debemos cultivar la autonomía y la humildad intelectual. Éstas son virtudes intelectuales que nos permiten, en parte, lidiar con la mezcla tóxica que esta tecnología y nuestra psicología generan. La autonomía intelectual consiste en la disposición y habilidad de pensar críticamente por uno mismo. En particular, le permite a uno monitorear y evaluar las creencias y argumentos de otros y de uno mismo, y así entonces convertirse en un agente epistémico responsable que no puede ser acusado de una ingenua credulidad. Como hemos dicho, en la Web

hay mucha data que es intencionadamente engañosa o simplemente falsa. Entonces, ¿qué podría ser más útil que desarrollar un filtro intelectual para discriminar las noticias falsas de las legítimas?

Esta virtud es doblemente importante hoy en día si uno desea diferenciar lo verdadero de lo falso, dado que uno no solo vive en una sociedad digitalizada sino también hiperespecializada. Para apreciar esto, considere nuestra dependencia en la pericia de otros. Éste es un aspecto ubicuo de la vida moderna. Vivimos en sociedades con conocimientos hiperespecializados, que dividen el trabajo epistémico entre sus miembros. Los seres humanos no somos máquinas de saber aisladas. No tenemos todos que ser doctores, abogados, científicos, etc. De hecho, nadie puede poseer todos estos conocimientos (o muchos de ellos). Si necesito saber qué le pasa a la pierna que me duele, iré del doctor y podré saber eso como también lo que debo hacer para revertir la situación. No necesito, ni puedo, saber todo. Cada uno tiene un rol en la división del trabajo epistémico y cada uno de nosotros puede, en principio, contar con el otro.

Pero esta dependencia epistémica requiere algún tipo de vigilancia o monitoreo con respecto a la selección de fuentes (por ejemplo, un supuesto experto) y la aceptación del contenido testificado dado que la gente puede engañarnos (por ejemplo, por interés) o puede estar equivocada (por ejemplo, por mera falibilidad o por incompetencia no reconocida sobre el asunto en cuestión). Por supuesto esta vigilancia epistémica tiene sus complicaciones cuando lidiamos como legos con supuestos expertos dada nuestra ignorancia sobre el asunto (y en particular cuando los expertos están en desacuerdo). Sin embargo, aprender a monitorear testimonios es algo mucho más manejable de aprender que los contenidos de los distintos campos de pericia. Y la autonomía intelectual promueve la efectiva vigilancia de fuentes y contenidos en todos los ámbitos. Además, lo anterior implica que cada uno de nosotros es ignorante (o tiene muy poco conocimiento en relación a los expertos) acerca



de diversos campos del conocimiento.

Para miembros de sociedades con conocimientos hiperespecializados, la ignorancia es inevitable. Esta ignorancia no es negativa dada la mencionada división del trabajo epistémico (en donde, como hemos visto, podemos contar con otros para el conocimiento que no poseemos) y nuestra capacidad para explotar tal división de trabajo (lo cual, como hemos visto, incluye ser intelectualmente autónomo). Entonces es importante reconocer que otros pueden saber más que uno sobre ciertos temas: particularmente, reconocer nuestra ignorancia y ser capaces de depender epistémicamente de ellos.

Pero el ignorante a menudo carece de la capacidad de reconocer su ignorancia. Esto puede generar en el ignorante una ilusión en relación al conocimiento que posee: una ilusión de conocimiento. Principalmente, el ignorante puede poseer una peligrosa sobreconfianza. Por supuesto que todos, como hemos visto, sufrimos del sesgo de la sobreconfianza. Pero dejando de lado esta tendencia general, el efecto Dunning-Kruger sugiere que más igno-

rante uno es, más confiado está de no serlo [11].

La ignorancia le otorga a uno una confianza engañosa que le roba la posibilidad de darse cuenta de su propia ignorancia. Éste es un fenómeno ampliamente observado y que se da justamente por la división del trabajo epistémico dado que la gente tiende a confundir lo que otros saben con lo que ellos saben [12]. De hecho, hay una versión digital de esta ilusión que es particularmente preocupante dada la amplia mediación de estas tecnologías en nuestras vidas. La gente tiende a sufrir esta ilusión de conocimiento sobre un tema cualquiera luego de haber buscado información en la Web por, inclusive, unos pocos minutos y sobre temas no relacionados [13].

Dado lo anterior, el peor enemigo del conocimiento no es la ignorancia, sino la ilusión de saber (creer equivocadamente que uno sabe). Ya que uno pierde la posibilidad de aprovechar la división de trabajo epistémico que está ahí justamente para ayudarnos a superar nuestra limitación de conocimiento. Y para aprovecharla y luchar contra estas ilusiones (incluyendo

las versiones digitales), un carácter intelectual humilde también parece necesario. Por lo tanto, la humildad intelectual también es doblemente importante en una sociedad digitalizada e hiperespecializada.

Esta humildad intelectual consiste en un conjunto de actitudes y disposiciones que debemos tomar hacia nosotros mismos y otros. Por ejemplo, reconocer que somos falibles: tener presente que probablemente no sabemos tanto como creemos y que probablemente nuestros sesgos cognitivos y prejuicios sociales pudieron haber afectado lo que creemos. También requiere que uno esté abierto a cambiar su posición según la evidencia y los argumentos lo sugieran. Esto no es solo estar abierto al cambio de opinión, sino al cambio dado el trabajo epistémico de otros (incluyendo el que se encuentra en la Web). Esta humildad puede ser entendida como un tipo de administración de confianza (de las creencias y capacidades epistémicas de uno) que nos permite aprovechar epistémicamente a otros. De ahí su importancia en un mundo hiperespecializado.

Por otra parte, esta humildad intelectual es también importante debido a que vivimos en un mundo digital dado que puede ayudarnos a contrarrestar la anteriormente mencionada arrogancia epistémica. En particular, esta virtud nos permite abrimos a las opiniones y argumentos de otros, incrementando así nuestras chances de detectar errores en nuestro pensamiento y de lidiar con la segregación epistémica que la tecnología de personalización promueve. Además, dado que la autonomía intelectual nos permite, como hemos visto, discriminar la buena de la mala argumentación (además de las buenas de las malas fuentes de información, etc.), es más

difícil aún intentar justificar paternalísticamente el tipo de intolerancia que, increíblemente, también encontramos en universidades.

Por lo tanto, si queremos diferenciar lo verdadero de lo falso en una sociedad digitalizada e hiperespecializada, debemos cultivar estas virtudes intelectuales. Pero esto no es fácil de realizar. Como nuestros sesgos lo sugieren, pensar críticamente y ser abierto de mente no surgen naturalmente (como el sexo o el chismear)—contrario en parte a los que algunos creen Aristóteles pensaba en la *Metafísica*: que cada uno de nosotros por naturaleza desea

saber, como si tuviéramos una curiosidad insaciable y un entusiasmo por aprender. Pero eso es muy cuestionable y mucha gente que ha enseñado durante muchos años probablemente diría que es obviamente falso. Una interpretación más plausible de Aristóteles (después de todo era una persona muy perspicaz) sería que no nos gusta estar equivocados. Deseamos saber en la medida en que deseamos evitar ser engañados y estar errados. Nos disturba descubrir que estamos equivocados acerca de algo. Si es esto lo que Aristóteles tenía en mente, entonces parece que estaba en lo correcto y es fácil entender el porqué: nuestras accio-



nes están basadas en gran parte en lo que pensamos es el caso y si estamos equivocados, éstas muy probablemente no serán exitosas.

Entonces, si deseamos no estar equivocados y dado que las virtudes que requerimos para evitar el error y el engaño en este mundo en el que vivimos no surgen naturalmente, es importante desarrollarlas para poder satisfacer ese deseo. Pero también parece que vivimos en un mundo donde la verdad, para muchos, ya no importa. Por eso puede existir un candidato a presidente en cuya compañía electoral 70% de las afirmaciones sean falsas, 2/3 del electorado entrevistado durante dicha campaña piense que no es una persona de confianza y sin embargo tal candidato gane las elecciones [14]. O un presidente que diga, en promedio, 7.6 afirmaciones falsas o engañosas al día y no pague ningún precio político por eso [15]. Entonces, se dice que vivimos en la “era de la postverdad” como si la verdad ya no importara. Pero la postverdad no implica que la verdad no existe o no importa sino más bien que los hechos están subordinados a nuestro punto de vista. La postverdad no respeta la verdad entendida como aquello que es independiente de nuestros sentimientos, creencias, etc. Uno, por supuesto, tiene derecho a su propia opinión, pero no a sus propios hechos. La realidad es una: hubo un ataque terrorista en Suecia o no lo hubo [16], es la ceremonia de investidura con mayor cantidad de gente en la historia del país o no lo es [17], uno gana el voto popular o no lo gana [18]. Nada de esto depende de lo que uno sienta o crea. Uno puede equivocadamente sentir que posee la verdad y también puede creer falsamente. La realidad es independiente de nuestra psicológica y es peligroso ignorar la realidad para acomodar nuestra psicología. El éxito de nuestros planes y proyectos dependen de la realidad y el peligro de la postverdad es el riesgo de distanciarse de la realidad.

Los sesgos y efectos psicológicos mencionados anteriormente están obviamente relacionados al fenómeno de la postverdad. Ellos son los precursores de la postverdad y cuando la tecnología de la información potencia esas deficiencias cognitivas, y de esa manera polariza las opiniones y fragmenta la población según sus opiniones, la postverdad surge. Es así como la mezcla de tecnología digital y psicología humana pueden crear un medio hostil para el debate público (no manipulativo). No solo por la segregación que promueven sino también por la erosión del concepto de verdad. Este tipo de debate típicamente busca dilucidar la verdad y supone la idea de una verdad común que responde a la realidad y que entre las distintas partes con sus distintas opiniones tienen mejores chances de alcanzar (ya que éstas pueden introducir razones que uno aisladamente no consideraría y así incrementar la probabilidad de que posiciones erróneas sean corregidas). Este debate sería un ejercicio fútil si la verdad respondiera a nuestra psicología en vez de la realidad. Es así entonces como nuestras interacciones con ciertas tecnologías digitales debilitan aún más nuestras chances de alcanzar la verdad.

DEMOCRACIA Y CIUDADANÍA

En todos los ámbitos de nuestras vidas es importante regular adecuadamente nuestras creencias para que nuestros planes y proyectos tiendan a ser exitosos. Esta regulación es también importante debido a que la democracia requiere que sus ciudadanos sean competentes: mínimamente, que estén bien informados y sean capaces de formar opiniones razonables. Sin embargo, hemos visto que nuestra interacción con ciertas tecnologías digitales de la información está cambiando radicalmente la manera en que nos informamos y formamos nuestras creencias. Éstas promueven, exacerbando y perpetúan deficiencias cognitivas que

nos alejan de la verdad y que crean un entorno donde la verdad no es respetada. El negativo impacto epistémico que esto tiene en la competencia ciudadana requerida para la democracia amenaza la posibilidad de un “gobierno del pueblo”. Si nada puede hacerse para satisfacer esa competencia mínima, tendremos que buscar alguna otra forma de gobierno (por ejemplo, como Platón nos sugiere, una epistocracia).

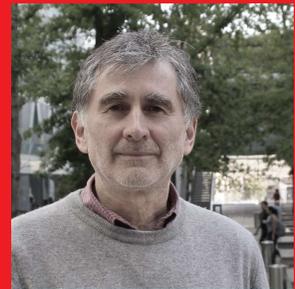
La democracia, por supuesto, no requiere solo esta competencia epistémica. Sin embargo, ésta es necesaria y sin ella no hay democracia verdadera. Pero así como nuestras interacciones con estas tecnologías pueden promover la desinformación, la polarización y la segregación, también pueden, como hemos visto, en conjunto con un determinado carácter intelectual, promover la verdad. Por lo tanto, si la democracia no ha de morir en nuestras propias manos, debemos por lo menos fomentar el sano y fructífero desarrollo de ciertas virtudes intelectuales, como la autonomía y la humildad intelectual. Éstas, como hemos visto, nos ayudarán a responder a los desafíos creados por nuestra tecnología y tomar ventaja de la hiperespecialización. En particular, estas virtudes le ayudarán al consumidor de noticias online a diversificar sus noticias (y así intentar destruir su burbuja digital), a realmente escuchar lo que dicen los que discrepan (y así intentar destruir su cámara de eco), y a aprender a juzgar los méritos de distintos ítems de información en la Web (y así intentar destruir las noticias falsas), entre otras cosas. Seguramente para lograr algunas de estas cosas, el consumidor de noticias online no solo tenga que adquirir estas virtudes intelectuales sino también pagar por los servicios online que utiliza (y así escaparle al modelo de publicidad que fomenta la personalización). Quizá nos demos cuenta a tiempo que ese no es un precio excesivo para salvar a la democracia. ■

REFERENCIAS

- [1] Brennan, J. 2016. *Against Democracy*. New Haven, Princeton University Press, p. 288.
- [2] Caplan, B. 2007. *The Myth of the Rational Voter*. New Haven, Princeton University Press, p. 276.
- [3] Delli Carpini, M. y Keeter, S. 1996. *What Americans Know about Politics and Why it Matters*. New Haven, Yale University Press, p. 397.
- [4] Doward, J., Catwalladr, C. y Gibbs, A. 2017. Watchdog to launch inquiry into misuse of data in politics. *The Guardian*, 4 marzo. <https://www.theguardian.com/technology/2017/mar/04/cambridge-analytics-data-brexit-trump>
- [5] Reuters Institute. 2018. *Digital News Report*. <http://www.digitalnewsreport.org/>
- [6] González, A. 2017. Bachelet alerta a los medios por noticias falsas que circulan en redes sociales. 7 abril. <https://www.biobiochile.cl/noticias/nacional/chile/2017/04/07/bachelet-alerta-a-los-medios-por-noticias-falsas-que-circulan-en-redes-sociales.html>
- [7] Mercier, H. 2017. Confirmation bias-myside bias. In Pohl, R. (ed.) *Cognitive Illusions*. London, Routledge, pp. 99-114.
- [8] Gorman, S. y Gorman, J. 2017. *Denying to the Grave*. Oxford, OUP, p. 328.
- [9] Hoffrage, U. 2017. Overconfidence. In Pohl, R. (ed.) *Cognitive Illusions*. London, Routledge, pp. 291-314.
- [10] Brookings Institution. 2017. Views among college students regarding the First Amendment: Results from a new survey. <https://www.brookings.edu/blog/fixgov/2017/09/18/views-among-college-students-regarding-the-first-amendment-results-from-a-new-survey/>
- [11] Kruger, J. y Dunning, D. 1999. Unskilled and Unaware of It: How Difficulties in Reconizing One's Own Incompetence Lead to Inflated Self-Assessment. *Journal of Personality and Social Psychology* 77(6):1121-1134.
- [12] Sloman, S. y Fernbach, P. 2017. *The Knowledge Illusion: Why we never think alone*. London, Macmillan, p. 296.
- [13] Fisher, M., Goddu, M. y Keil, F. 2015. Searching for Explanations: How the Internet Inflates Estimates of Internal Knowledge. *Journal of Experimental Psychology* 144(3):674-687.
- [14] Gibbs, N. 2017. When a President Can't Be Taken at His Word. *Time*, 23 Marzo. <http://time.com/4710615/donald-trump-truth-falsehoods/>
- [15] Kessler, G., Rizzo, S. y Kelly, M. 2018. President Trump has made 4,229 false or misleading claims in 558 days *The Washington Post*, 1 agosto. <https://www.washingtonpost.com/news/fact-checker/wp/2018/08/01/president-trump-has-made-4229-false-or-misleading-claims-in-558-days>
- [16] BBC. 2017. Trump: Look at what happened last night in Sweden. 19 febrero. <https://www.bbc.com/news/av/world-us-canada-39021599/trump-look-at-what-happened-last-night-in-sweden>
- [17] Porter, T. 2018. Sean Spicer admits he 'screwed up' by claiming trump had biggest-ever inauguration audience. *Newsweek*, 24 julio. <https://www.newsweek.com/sean-spicer-admits-he-screwed-claiming-trump-inauguration-crowd-biggest-ever-1039439>
- [18] Relman, E. 2017. Nearly half of Republicans believe Trump won the popular vote in November. *Business Insider*, 26 julio. <https://www.businessinsider.com/repUBLICANS-think-trump-won-popular-vote-2017-7>



RANKINGS DE PERSONAS: EL CASO DEL SISTEMA DE CRÉDITO SOCIAL CHINO



CLAUDIO GUTIÉRREZ

Profesor Titular del Departamento de Ciencias de la Computación de la Universidad de Chile. Investigador Senior del Instituto Milenio Fundamentos de los Datos. Ph.D. Computer Science, Wesleyan University.

Líneas de investigación: fundamentos de la computación, lógica aplicada a la computación, bases de datos, semántica de la Web, máquinas sociales.
cgutierr@dcc.uchile.cl

Según la RAE, un ranking es "una clasificación de mayor a menor, útil para establecer criterios de valoración". El tema que nos ocupará son rankings de personas, esto es, "el acto de resumir nuestro juicio sobre el desempeño de una persona en un único, holístico número o nota"¹, rankings -particularmente aquellos producidos por algoritmos- que se han transformado en ubicuos hoy día.

RANKINGS DE PERSONAS

Aunque no todos seamos conscientes, estamos rodeados de rankings de personas. La Asociación de Profesionales del Tenis (ATP) mantiene un ranking de sus jugadores. En ajedrez ocurre algo similar con el ELO. La PSU puede considerarse un ranking anual de los estudiantes que egresan de la enseñanza media. En la academia contemporánea, existen rankings de personas por sus artículos o citas. De manera similar, muchas empresas recogen datos personales para usarlos en rankings implícitos, muchas veces no públicos. Por ejemplo las aerolíneas tienen rankings de millas de vuelo por pasajero. El mundo financiero clasifica a sus clientes en rankings que incluyen facetas como sueldo, estabilidad laboral, capacidad de pago, familia, etc. En Chile existe el infame reporte DICOM, con el que una empresa privada clasifica a personas como confiables o no financieramente. Las isapres también manejan rankings para clasificar a sus clientes y cobrarles acordeamente.

Los ranking de personas clasifican a éstas por alguna "cualidad", que como toda faceta humana, las más de las veces no tiene métrica precisa. En algunos deportes como el tenis se facilita porque se reduce al desempeño en ciertos torneos. Pero en fútbol, por ejemplo, no es evidente cómo hacer un ranking de los mejores jugadores. Estos rankings se tornan éticamente más oscuros en otros ámbitos. Por ejemplo, en algunas empresas la dimensión más relevante es el dinero, y luego las personas serán ordenadas según eso. Así un "buen" paciente para una isapre es alguien cuya salud cuesta poco, y eso reducido a parámetros medibles será joven, hombre, sin antecedentes,

etc. Para la empresa Thomson Reuters cuyo negocio son las publicaciones científicas, "buen" investigador es quien publica más y es más citado en su sistema, esto es, quien le produce más ganancias. Estos criterios no son los mismos que los que la comunidad médica considera un buen paciente o los que la comunidad científica considera un buen investigador. No es difícil vislumbrar los gigantes desafíos éticos que los rankings de personas involucran.

Con el advenimiento del big data, los rankings de personas han comenzado a ser muy comunes, más "eficientes" (para los objetivos de sus promotores) y están en algunas partes de manera casi invisible debido al desarrollo de técnicas computacionales de captura y procesamiento de datos implícitos. Las grandes compañías de datos como Amazon, Facebook, Twitter, Google, eBay, etc. han convertido este tema en su negocio principal. Todas cuentan con (y basan su negocio en) sofisticados rankings de personas en diversas áreas para diferentes objetivos basados en sus datos: páginas más visitadas, personas con más "amigos", personas más populares, etc. Entre los más conocidos está PageRank de Google y los sistemas de reputación de eBay, Amazon, etc. Ellos usan sofisticadas teorías y técnicas estadísticas, de ciencias sociales y políticas, métricas científicas, de computación. Entre ellas destacan sistemas de ranking, agregación de preferencias, *social choice theory*, teoría de votaciones, teoría de juegos, agregación de reglas, análisis de citas, teoría de grafos, etc.

Entre los mayores desafíos técnicos que enfrenta un software que elabore rankings de personas están: qué datos incluir que representen ojalá la mayor cantidad de facetas y cómo obtenerlos; ser capaz de combinar y ponderar adecuadamente esa inmensidad de datos (para calibrar esto se requieren técnicas de aprendizaje de máquina); que la persona no esté consciente de los datos que está entregando (tanto para facilitar su captura como para evitar la manipulación del ranking); y el que el sistema vaya mejorando y adaptándose con el paso del tiempo. El dominio actual de estas técnicas (sensores y capacidad de almacenamiento y procesamiento masivo; redes sociales; apren-

dizaje de máquinas) es lo que ha disparado los usos de estos rankings hoy.

RANKING UNIVERSALES: EL SISTEMA DE CRÉDITO SOCIAL CHINO

Si uno extrapola los casos mencionados antes, no pareciera descabellado a primera vista preguntarse si no pudiera existir un ranking "universal" para clasificar humanos. En lo que sigue nos concentraremos en el que es probablemente el experimento social más radical y audaz de los últimas décadas: el ranking de ciudadanos en China conocido como "Sistema de Crédito Social" (SCS en adelante).

China es una sociedad de más de 1.400 millones de habitantes, la gran mayoría recién incorporándose a formas de vida moderna. Por ello, uno de los grandes problemas para el gobierno chino es la falta de mecanismos para evaluar los niveles de confianza de sus ciudadanos, particularmente para el sistema financiero, pero también para el sistema social y cultural en general. Para abordar ese problema, el año 2014 el gobierno chino comenzó a delinear la construcción de un sistema de crédito social (SCS), un sistema a nivel nacional que determina y asigna niveles de confianza a los ciudadanos. La idea, dicen sus promotores, es construir y fortalecer una cultura de sinceridad: "[El Sistema] forjará un ambiente de opinión pública donde mantener la confianza será glorioso. Fortalecerá la sinceridad en los asuntos de gobierno, la sinceridad comercial, la sinceridad social y la construcción de credibilidad judicial". El Sistema permitirá premiar los comportamientos que el gobierno considere responsables (financiera, económica y sociopolíticamente) y sancionar a quienes se alejen de esas políticas. Desde el punto de vista computacional, la gran novedad es la combinación de fuentes "tradicionales" de datos, como financieros, criminales y gubernamentales en general, con el uso masivo de sensores digitales y artefactos (Internet de las cosas, teléfonos móviles, etc.) que puedan recolectar de manera continua y evaluar datos de comportamiento en tiempo real.

1. P. Elbow. Es importante hacer notar que un ranking no es una evaluación. Una evaluación (por parte de un profesor a un alumno por ejemplo) es el acto de expresar el juicio que el profesor se hace del desempeño del estudiante apuntando a las fortalezas y debilidades de diferentes aspectos en diferentes dimensiones de su quehacer. Esto lo distancia radicalmente del ranking: no reduce el desempeño a una comparación con otros, y menos a un orden lineal.

Este ranking de personas, que servirá para diversos efectos, es voluntario por ahora, pero en 2020 comenzará a ser obligatorio. Actualmente está en período de pruebas experimentales en algunos gobiernos provinciales y algunas compañías. Dos de ellas, que manejan muchos datos, ya tienen dos proyectos: *China Rapid Finance* por Tencet (el Facebook chino) y *Sesame Credit*, de Ant Financial Services Group (AFSG), compañía afiliada de Alibaba. Hoy hay mucha gente ingresando al sistema. ¿Por qué? Por las mismas razones que la gente ingresa a los concursos, por las que espera obtener buenas puntuaciones en Uber, o muchos likes en Facebook, etc. Esperan posibles beneficios si están bien en el ranking.

Los factores que el sistema de Alibaba hoy considera para construir su algoritmo (cuya estructura no es pública) son: historia crediticia; capacidad de cumplimiento; características personales; comportamiento y preferencias; y relaciones interpersonales. En principio se ha anunciado que solo asuntos positivos contarán. No es difícil imaginar la gigantesca cantidad de variables que involucra cada uno de esos factores.

DESAFÍOS Y CONTROVERSIAS

El Sistema, como es de esperarse, ha generado bastante controversia y muchas inquietudes y críticas.

Por un lado, están aquellos que ven un perfecto sistema orwelliano donde los ciudadanos pasarán a ser marionetas del gobierno. Otros desde fuera de China reclaman que eso producirá una competencia desleal entre empresas chinas y extranjeras. En la misma sociedad china y entre sus promotores se ha planteado que la infraestructura gubernamental es lenta y está retrasada para asumir un proyecto de esa magnitud; los agentes del mundo financiero chino señalan los peligros de combinar factores sociales y económicos en un mismo índice; y finalmente aparecen preocupaciones sobre la seguridad en el sentido de evitar intentos de engañar el sistema o posibles hackeos del sistema.

Desde una posición más positiva, hay quienes observan este experimento social intentando entender sus posibles desarrollos. Se señala también que el tema de los rankings de personas es algo que llegó para quedarse, y que este experimento a nivel gubernamental no hace más que sincerar esa situación. En particular, expondrá muy claramente los alcances y límites de los usos de los datos personales. En la sociedad china aún no hay conciencia generalizada sobre sus alcances, aunque en los proyectos parciales se ha mostrado mucho entusiasmo de la gente por incorporarse. El gobierno chino ha planteado que el SCS pudiera permitir que "la confianza roja en todos los rincones bajo el cielo haciendo difícil que aquellos desacreditados engañen a otros". ¿No sería ideal tener un sensor que nos dijera cuán confiable es la persona que tenemos al lado?

Como vemos, estamos ante lo que significa un acontecimiento: algo cuyos alcances y derivaciones no están previstos ni definidos de antemano. Particularmente, los usos que el sistema pueda hacer de esa información para direccionar comportamientos sociales. La pesadilla de un sistema así podemos intuirlo al conocer cuán sesgados son los sistemas construidos por los humanos hoy día. Y qué se puede hacer con información como esa sin control social ni público (el caso de Cambridge Analytica es un buen ejemplo). Pero por el lado positivo, la comunidad tendría una herramienta para construir una sociedad de confianzas, para conocer cuán confiable socialmente es quien se nos aproxima, cuán confiable es quien encuentro en una organización, etc. La ciencia ficción apunta a los extremos, de la pesadilla y el ideal. Si algo tiene de futurista el experimento del SCS es que avanza decididamente, de manera "abierto", pública (esto es, no privada) en un terreno que es escabroso y sujeto a la fortaleza y fragilidad de lo humano². Y de aquí sale una lección para quienes trabajamos en computación: asumir y considerar las derivaciones de nuestros desarrollos técnicos.

Se diga lo que se diga, el SCS chino es una de las iniciativas sociales que marcarán los próximos años. Y de alguna manera es sincerar

una tendencia que hoy día vivimos inconscientemente cada vez que cliqueamos en Google, visitamos Amazon, marcamos un tweet o ingresamos a Facebook. ¿Es positivo o negativo? Como escribe Kevin Kelly en *Lo Inevitable*, nuestra elección ahora es si queremos que la vigilancia sea secreta, un *panopticon* de una dirección, o una covigilancia mutua transparente que involucra observar a los observadores.

Una vez más queda claro que los asuntos más relevantes de la construcción de futuro no solo son tecnológicos o económicos, sino éticos y políticos. Como en todo lo social, mucho dependerá de todos nosotros hacia dónde se encaminen estas iniciativas. ■

REFERENCIAS

- Rachel Botsman. "Big data meets Big Brother as China moves to rate its citizens", *Wired*, 21 Oct. 2017.
- Peter Elbow. "Ranking, Evaluating, Liking: Sorting Out Three Forms of Judgment." *College English*. Vol. 55, No. 2 (Feb., 1993), pp. 187-206.
- Huang Jian-xin, Wu Yan. "Some problems in establishment of social credit system in China", *Journal of Harbin Institute of Technology (Social Sciences Edition)*, 2002-03.
- Mareike Ohlberg, Shazeda Ahmed, Bertram Lang. "Central Planning, Local Experiments. The complex implementation of China's Social Credit System." <https://www.merics.org/en/china-monitor/content/5071#6636>
- Kevin Kelly. *The Inevitable*. Viking Press, 2016.
- Felix Brandt, Vincent Conitzer, Ulle Endriss, Jerome Lang, Ariel D. Procaccia. Introduction to Computational Social Choice. En: *Handbook of Computational Social Choice*, Cambridge Univ. Press, 2015.

2. No estamos con esto sosteniendo opinión positiva o negativa sobre el sistema de Gobierno chino. Estamos diciendo que plantear abiertamente que se construirá un ranking social con esas características difiere fundamentalmente de rankings de empresas privadas o peor, de experimentos secretos como los de Cambridge Analytics. De cualquier forma, el devenir de un sistema así dependerá del control de los ciudadanos sobre el gobierno.

¿QUÉ ES GDPR?



**LUIS
ARANCIBIA**

Abogado. Actualmente se desempeña en la Administración del Dominio .CL (NIC Chile) en el área de gestión de resolución de controversias por nombres de dominio .CL, entre otras materias. Miembro del Directorio de la Asociación de ccTLD's de América Latina y el Caribe.

lam@nic.cl

INTRODUCCIÓN

El 2018 podrá ser recordado como un año clave en la protección de datos en todo el mundo. A partir de enero de 2018, más de 100 países han promulgado una legislación integral de protección de datos, y alrededor de 40 países están en proceso de establecer dichas leyes¹ y ². Nuestro país no ha estado ajeno a dicho proceso. El 15 de mayo de 2018 se aprobó una reforma a la Constitución, la cual estableció explícitamente el derecho a la protección de datos personales como un derecho de rango constitucional, al tiempo que se encuentra en actual discusión el proyecto de ley para una nueva ley de datos personales, que pretende sustituir el régimen establecido por la ley 19.628 sobre Protección de Datos de Carácter Personal de agosto de 1999.

En este contexto se despliega -si es que acaso no es el gatillador de dicho proceso- el Reglamento de Protección de Datos de la Unión Europea, también conocido como GDPR, el cual podría ciertamente convertirse en un modelo para el resto del mundo, ya que mu-

chas empresas globales prestan servicio a usuarios en y hacia la Unión Europea. De este modo, dichas empresas y otras organizaciones tendrán que adaptarse a estas reglamentaciones de todos modos, y podría tener sentido adoptar los principios de la norma europea de privacidad digital en todo el mundo³.

El GDPR es, sin duda, el desarrollo regulatorio más importante en materia de protección de datos personales vigente hasta la fecha. Basado en un mecanismo detallado y protector, está influyendo en el uso de los datos personales y a muchas legislaciones en todo el mundo. Entendido correctamente, el GDPR alentará a las empresas a desarrollar marcos precisos y explícitos de gestión de la información, a utilizar los datos personales en base a una práctica que las obligará a mantener a los titulares de los datos informados sobre la toma de decisiones respecto de ellos. Para alcanzar estos objetivos, el GDPR utiliza criterios y herramientas jurídicas y procedimentales, que serán los factores estructurales que facilitarán la comprobación de violaciones, a la vez que pondrá en tela de juicio a algunos modelos de negocio concebidos en base al uso intensivo de información de personas.

ANTECEDENTES

GDPR es la abreviatura que refiere al General Data Protection Regulation, conocido también como el Reglamento de Protección de Datos de la Unión Europea que entró en vigencia el 25 de mayo de 2018.

El GDPR es la nueva regulación europea para la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de los mismos, el cual fue adoptado como Reglamento (UE) 2016/679 del Parlamento Europeo y el Consejo de la Unión Europea el 27 de abril de 2016⁴ y ⁵.

El Reglamento es directamente aplicable en todos los estados miembros de la Unión a partir del 25 de mayo de 2018 y está diseñado para reforzar los mecanismos de protección del derecho fundamental de las personas físicas en relación con el tratamiento de sus datos personales⁶. En base a ello, el Reglamento reconoce que los principios y normas relativos a la protección de las personas físicas, en lo que respecta al tratamiento de sus datos

1. Banisar, David. *National Comprehensive Data Protection/Privacy Laws and Bills 2018* (January 25, 2018), disponible en: <https://ssrn.com/abstract=1951416> or <http://dx.doi.org/10.2139/ssrn.1951416>

2. Algunos países tienen leyes de privacidad que se aplican a ciertas áreas, por ejemplo, para niños o registros financieros, pero no tienen una ley integral sobre protección de datos. En los países donde no existe un marco integral de protección de datos, la protección de datos está regulada por leyes sectoriales. Por ejemplo, a pesar de ser un líder temprano en el campo de la protección de datos, la Ley de Privacidad de Estados Unidos de 1974 se aplica únicamente al Gobierno Federal y las leyes posteriores se aplican a sectores o grupos específicos de personas (por ejemplo, la Ley de Protección de la Privacidad en Línea de los Niños). Hasta la fecha no existe una ley integral de protección de datos. Sin embargo, dentro del proceso descrito cabe mencionar la entrada en vigencia de la Ley de Protección de Datos Personales en Posesión de Sujetos Obligados para el Estado de Baja California, del 18 de agosto de 2017. Ver: A Guide for Policy Engagement on Data Protection. PART 1: Data Protection, Explained, disponible en <https://www.privacyinternational.org/>

3. Una crítica a esta posibilidad, véase Chakravorti, Bashar: *The rest of the world can't free ride on GDPR*, disponible en <https://www.weforum.org/agenda/2018/05/why-the-rest-of-the-world-can-t-free-ride-on-europe-s-gdpr-rules>

4. El texto oficial puede obtenerse en: <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX%3A32016R0679>

5. Junto al GDPR, la reforma consideró también la aprobación de la Directiva (UE) 2016/680 del Parlamento Europeo y del Consejo de 27 de abril de 2016 relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales por parte de las autoridades competentes para fines de prevención, investigación, detección o enjuiciamiento de infracciones penales o de ejecución de sanciones penales, y a la libre circulación de dichos datos, por la que se deroga la Decisión Marco 2008/977/JAI del Consejo.

6. De acuerdo al sistema de fuentes del Derecho de la Unión Europea, un Reglamento tiene alcance general, es obligatorio en todas sus partes y directamente aplicable en cada estado miembro. Los destinatarios, esto es, personas, estados miembros e instituciones de la Unión, deberán acatar los Reglamentos en su totalidad como "ley de la Comunidad". Un Reglamento es directamente aplicable en todos los estados miembros, sin necesidad de un acto nacional de adopción. Desde su entrada en vigor, se impone en todos los ordenamientos jurídicos nacionales. Una Directiva, en cambio, obliga al estado miembro destinatario en cuanto al resultado que debe conseguirse y son ellos los que eligen la forma y los medios para la adopción de sus términos. Para que los objetivos contemplados en la Directiva tengan efecto para los ciudadanos es preciso que el legislador nacional proceda a un acto de transposición o "medida ejecutiva nacional", mediante el cual el Derecho nacional se adapta a los objetivos determinados en la misma. Borchardt, Klaus-Dieter. *El ABC del Derecho de la Unión Europea*. Oficina de Publicaciones de la Unión Europea. 2011, p. 93.

de carácter personal deben respetar sus libertades y derechos fundamentales, cualquiera que sea su nacionalidad o residencia.

El GDPR deroga la Directiva 95/46/CE del Parlamento Europeo y del Consejo, en función de armonizar la protección de los derechos y las libertades fundamentales de las personas físicas en relación con las actividades de tratamiento de datos de carácter personal y garantizar la libre circulación de dichos datos entre los estados miembros. Tal como sostiene uno de los considerandos del Reglamento⁷, aunque los objetivos y principios de la Directiva 95/46/CE siguen siendo válidos, ello no ha impedido que la protección de los datos en el territorio de la Unión se aplique de manera fragmentada ni ha evitado la inseguridad jurídica o una percepción generalizada entre la opinión pública de que existen riesgos importantes para la protección de las personas físicas, en particular en relación a sus actividades en línea. Las diferencias en el nivel de protección en los estados miembros respecto del derecho a la protección de los datos de carácter personal, pueden impedir la libre circulación de éstos en la Unión. Estas diferencias pueden convertirse, por lo tanto, en un obstáculo al ejercicio de las actividades económicas, deteriorar la competitividad e impedir que las autoridades cumplan las funciones que les incumben en virtud del Derecho de la Unión. Esta diferencia en los niveles de protección se debió a la existencia de divergencias en la ejecución y aplicación de la Directiva 95/46/CE⁸.

El nuevo estatuto se establece para garantizar un nivel uniforme y elevado de protección de las personas físicas y eliminar los obstáculos a la circulación de datos personales dentro de la Unión, en consideración a la determinación

de un nivel de protección equivalente, coherente y homogéneo en todos los estados miembros, sin perjuicio de que, en lo que respecta al tratamiento de datos personales para el cumplimiento de una obligación legal, para el cumplimiento de una misión realizada en interés público o en el ejercicio de poderes públicos conferidos al responsable del tratamiento, los estados deben estar facultados para mantener o adoptar disposiciones nacionales a fin de especificar en mayor grado la aplicación de las normas del Reglamento. Junto con la normativa general y horizontal sobre protección de datos por la que se aplica la Directiva 95/46/CE, los estados miembros cuentan con distintas normas sectoriales en ámbitos que precisan disposiciones más particulares. El Reglamento reconoce también un margen de maniobra para que los estados miembros especifiquen sus normas, inclusive para el tratamiento de categorías especiales de datos personales («datos sensibles»). En este sentido, el Reglamento no excluye el Derecho de los estados miembros que determina las circunstancias relativas a situaciones especiales de tratamiento, incluida la indicación pormenorizada de las condiciones en las que el tratamiento de datos personales es lícito.

Al tener un impacto directo en todos los estados miembros, el GDPR hará que muchos de ellos deban modificar sus leyes nacionales para regular aspectos tales como la posición de la Autoridad de Protección de Datos, las reglamentaciones sectoriales, las normas de transición o las implementaciones de requisitos adicionales en los casos en que el GDPR otorga discreción. Ya se han publicado los primeros proyectos de leyes nacionales con los cambios legislativos necesarios, por ejemplo, en Alemania, los Países Bajos y Polonia.

PRINCIPALES CONTENIDOS DEL GDPR

El Reglamento diseña un modelo de políticas de la Unión Europea en materia de protección de datos personales. Aunque se mantienen los principios de la Directiva 95/46/CE, se han propuesto cambios importantes a las políticas imperantes, que crean un estado de cosas nuevo en el estatuto del tratamiento de los datos personales, con un alto impacto en muchas industrias europeas y globales.

A continuación se expone un resumen de los principales aspectos del GDPR.

(A) MAYOR ALCANCE TERRITORIAL (APLICABILIDAD EXTRATERRITORIAL)

El Reglamento aplica a todo el procesamiento de datos personales de personas físicas que permanezcan en la Unión, independientemente de la ubicación de la empresa. En este aspecto el propósito de la regulación es clara: el GDPR se aplicará al procesamiento de datos personales que realicen controladores y procesadores en la Unión Europea, sin importar si el proceso de tratamiento de los datos tiene lugar o no en este espacio. Asimismo, también se aplicará al procesamiento de datos personales de personas en la UE por un controlador o procesador no establecido en la Unión, donde las actividades de cualquiera de aquellos se relacionan con la oferta de bienes o servicios a interesados de la UE (independientemente de si se requiere o no pago) o el control de su comportamiento en la medida que tenga lugar dentro de la UE⁹.

7. Véase considerando 1º a 12º del texto del Reglamento.

8. La versión de la Directiva es accesible en: <http://www.wipo.int/wipolex/es/details.jsp?id=13580>. Una explicación de sus principales aspectos puede consultarse en Cerda, Alberto. Determinación informativa y leyes de protección de datos. Revista Chilena de Derecho Informático. N°3, Diciembre, 2003. Disponible en http://web.uchile.cl/vignette/derechoinformatico/CDA/der_informatico_completo/0,1492,SCID%253D14331%2526SID%253D507,00.html

9. El Reglamento distingue entre ámbito de aplicación material y territorial. Sobre el Ámbito de Aplicación Material del Reglamento, véase nuestro artículo <http://www.laleyaldia.cl/?p=5308>



(B) RÉGIMEN DE SANCIONES

Las organizaciones que infrinjan el GDPR, podrán ser multadas con hasta el 4% de la facturación global anual o €20 millones (lo que sea mayor), de acuerdo a lo señalado en el artículo 83, número 5. Ésta es la multa máxima que se puede imponer por las infracciones más graves, como por ejemplo, el no haber obtenido el suficiente consentimiento del interesado para procesar sus datos. Existe un esquema escalonado de multas. A su vez, es importante tener en cuenta que estas reglas se aplican tanto a los controladores como a los procesadores, lo que significa que las “nubes” no estarán exentas de la aplicación de GDPR.

(C) EL CONSENTIMIENTO

El GDPR fortalece las condiciones del consentimiento, el cual se somete un régimen de mayor rigor. Con los nuevos criterios, se terminará con el consentimiento tácito. En este sentido, se deberá entender por consentimiento del interesado a toda manifestación de voluntad libre, específica, informada e inequívoca por la cual el sujeto acepta, ya sea por una declaración o una clara acción afirmativa, el tratamiento de los datos que le conciernen. El consentimiento debe ser claro y distinguible de otros asuntos y proporcionado en una forma inteligible y de fácil acceso, usando un lenguaje claro y sencillo. El interesado

tendrá derecho a retirar el consentimiento en cualquier momento. En términos del GDPR, deberá ser tan fácil retirar el consentimiento como darlo.

(D) DERECHOS DE LOS INTERESADOS

d.1. Notificación de filtración de datos. Bajo el GDPR, la notificación por filtración de datos será obligatoria en todos los estados miembros cuando sea probable que esta violación de la seguridad de los datos personales entrañe un alto riesgo para los derechos y libertades de las personas físicas. En estos casos, el responsable del tratamiento comunicará la probable violación dentro de las 72 horas de



los datos personales que los conciernen se están procesando, dónde y con qué propósito. Además, el controlador deberá proporcionar una copia de los datos personales, sin cargo, en un formato electrónico.

d.3. Derecho de supresión o el derecho al olvido. Conocido como borrado de datos, el derecho a ser olvidado da la posibilidad al interesado a solicitar que el responsable del tratamiento elimine sus datos personales, cese la diseminación de ellos y, potencialmente, haga que terceros detengan el procesamiento de los mismos. Incluso, en determinados casos, el destinatario de la petición deberá proceder sin dilación indebida, por ejemplo cuando los datos personales ya no sean necesarios para los fines para los que fueron recogidos o tratados, o cuando ellos fueron tratados ilícitamente. Sin embargo, este derecho está limitado y no aplica cuando sea necesario conservar los datos por razones de interés público para fines de investigación científica, histórica o estadísticos, entre otros.

d.4. Derecho a la portabilidad de datos. Entendida como la facultad del interesado de recibir los datos personales que le conciernen, que previamente ha proporcionado, en un formato estructurado, de uso común y lectura mecánica, así como el derecho a transmitir esos datos a otro controlador, sin posibilidad de que el anterior detentador de los mismos pueda impedirlo.

(E) PRIVACIDAD POR DISEÑO

La privacidad por diseño implica la exigencia de la inclusión de la protección de datos desde el inicio del diseño de los sistemas, en lugar de ser una adición. Más concretamente, el responsable del tratamiento deberá aplicar las medidas técnicas y organizativas adecuadas

de manera efectiva para cumplir los requisitos del Reglamento y proteger los derechos de los interesados. El artículo 25 exige que los controladores retengan y procesen solo los datos absolutamente necesarios para el cumplimiento de sus funciones (minimización de datos), y que limiten el acceso a los datos personales a quienes necesitan representar el procesamiento.

(F) DELEGADOS DE PROTECCIÓN DE DATOS

Bajo las reglas de la Directiva, los controladores están obligados a notificar sus actividades de procesamiento de datos a las Autoridades de Protección de Datos locales, lo que, por ejemplo, para las multinacionales, puede ser una pesada carga burocrática, ya que la mayoría de los estados miembros tienen requisitos de notificación diferentes. Bajo el GDPR no será necesario enviar notificaciones de actividades de procesamiento de datos a cada Autoridad local, ni será un requisito notificar u obtener aprobación para transferencias. En su lugar, habrá requisitos internos de mantenimiento de registros y el nombramiento de un Oficial de Protección de Datos será obligatorio solo para aquellos controladores y procesadores cuyas actividades centrales consisten en operaciones de procesamiento que requieren un monitoreo regular y sistemático de datos a gran escala o de especiales categorías de datos o datos relacionados con condenas y delitos penales. Es importante destacar que el Delegado deberá nombrarse en función de las cualidades profesionales y, en particular, de los conocimientos especializados sobre legislación y prácticas de protección de datos, entre otros requisitos, expresados en el artículo 37 de GDPR. ■

haberla constatado por primera vez. Los procesadores de datos también deberán notificar a sus clientes, los controladores, "sin demora indebida" después de tomar conocimiento por primera vez de una violación de datos¹⁰.

d.2. Derecho de acceso. Consiste en el derecho de cualquier interesado para obtener la confirmación del controlador de los datos si

10. Estas reglas han sido puestas en práctica en el reciente incidente sufrido por British Airways, que reveló el jueves 6 de septiembre de 2018 que su sitio web y aplicación fueron víctimas del mayor hackeo que ha sufrido. El ataque es el primero en golpear a una compañía bajo las nuevas regulaciones del GDPR. La compañía, de acuerdo a los reportes de la prensa, habría notificado a la Oficina del Comisionado de Información y a los clientes dentro de las 72 horas obligatorias del Reglamento, sin perjuicio de estar bajo investigación sobre si hubo incumplimiento, lo que en caso de ser efectivo, podría implicar que la empresa sea penalizada si no toma todas las medidas necesarias para proteger los datos de los clientes.

Ver: <https://hipertextual.com/2018/09/british-airways-afectados-hackeo>

LA HAZAÑA DEL RECUENTO PARALELO DEL PLEBISCITO DE 1988





**JUAN
ÁLVAREZ
RUBIO**

Académico del Departamento de Ciencias de la Computación de la Universidad de Chile. Master of Mathematics (Computer Science), University of Waterloo. Ingeniero de Ejecución en Procesamiento de la Información, Universidad de Chile. Junto a su labor como docente, trabaja en reconstruir la historia de la computación en Chile.
jalvarez@dcc.uchile.cl

INTRODUCCIÓN

El 5 de octubre de 2018 se conmemoraron en Chile 30 años de la realización del plebiscito de 1988, en que la ciudadanía se pronunció por terminar con la dictadura que comenzó en 1973. En ese contexto, y como ha sido habitual en anteriores oportunidades, se recordaron solo algunos episodios como el rol y las acciones de algunos personajes y la “franja televisiva del NO”. Sin embargo, además de la gran influencia del movimiento social previo, hay una hazaña que es poco conocida y rara vez mencionada por los medios y los políticos y científicos sociales: el sistema socio-técnico de recuento paralelo de la votación que involucró a decenas de miles de anónimos voluntarios(as).

A comienzos de 1988 convergieron varias organizaciones de distintas tendencias políticas, que propiciaban elecciones libres como alternativa al plebiscito que daría continuidad al régimen, y constituyeron el “Comando por el NO”. El Comando definió tres líneas de trabajo: política, técnica (de información y análisis) e inscripción electoral (en el registro abierto en febrero de 1987). Esta última comisión se encargó también de diseñar un Sistema de Recuento Paralelo (SRP) al oficial con el propósito de evitar un probable manejo y distorsión de los resultados, tal como ocurrió en la consulta de 1978 y en el plebiscito de 1980.

EL SISTEMA DE RECUENTO PARALELO

El SRP fue diseñado considerando cuatro principios: confiabilidad (como alternativa al recuento del gobierno), velocidad (al menos comparable al sistema oficial), seguridad (para soportar hostigamientos de baja intensidad) y discriminación (basado en resultados por mesa). Por otra parte, se distinguieron tres restricciones principales: inexperiencia (para sistemas de esta

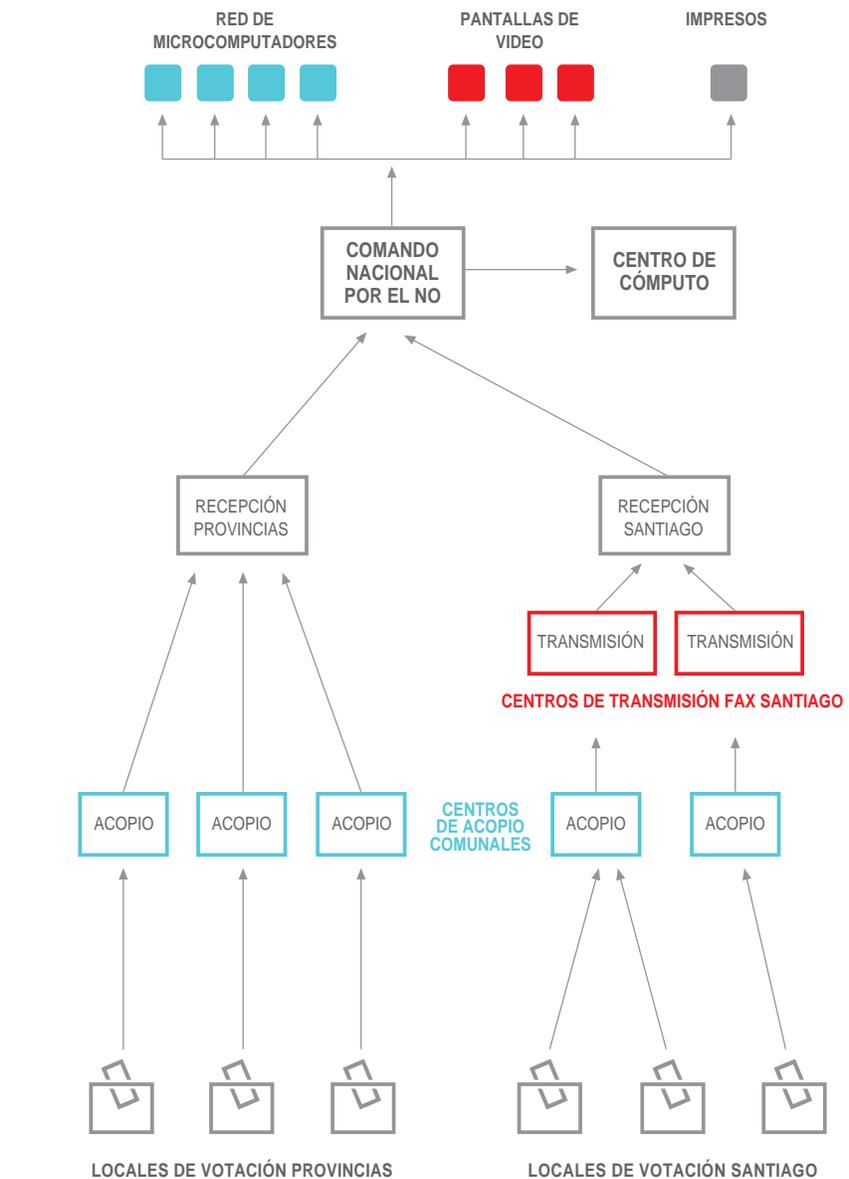


FIGURA 1
DISEÑO DEL SISTEMA DE RECUENTO PARALELO.

naturaleza), recursos (reducidos) y política (organización de miles de personas de distintas sensibilidades y militancias políticas) [1, 3].

El diseño del SRP, que fue realizado por un pequeño equipo de jóvenes ingenieros, comenzó en septiembre de 1987 y su versión final se definió a fines de junio de 1988 [2]. La arquitect-

tura piramidal del SRP comprendió las siguientes componentes [4]:

- Centros recolectores de la información de las mesas en los locales de votación (aproximadamente 1.100 centros para más de 22.000 mesas de votación).

- Centros de acopio y transmisión (por telefax) en las cabeceras de provincia (40).
- Centros de acopio comunales en Santiago (35) que proporcionaban la información a 7 centros de transmisión (por telefax).
- Un centro de transmisión oral para recibir información de locales de votación aislados.
- Centros de recepción: uno para información recibida de provincias y otro para Santiago.
- Centro de procesamiento (o de cómputo) en el Comando con un computador con 18 terminales para ingreso de información (y un centro de cómputo manual como respaldo).
- Centro de procesamiento alternativo en un lugar secreto.
- Conjunto de centros de despliegue de información.

A partir del diseño arquitectónico (**Figura 1**) se definieron cinco subsistemas [3]: Recolección de la Información (desde el recuento en las mesas hasta su comunicación al exterior del local de votación), Transmisión de la Información (desde los locales hacia los centros de transmisión y al computador central), Procesamiento de la Información (con un computador central y una red de microcomputadores para distribuir y desplegar los resultados) (**Figura 2**), Análisis y

Despliegue de la Información (evaluación estadística para proyectar resultados), Control (recuento manual, procedimientos de repliegue, respaldo computacional, respaldo en anillo de PCs, centro especial de control).

La instalación de los FAX se hizo dos semanas antes del plebiscito en solo tres días por parte de un equipo de estudiantes de ingeniería eléctrica. Por su parte, la red de microcomputadores fue instalada y operada por estudiantes de ingeniería en computación. La mayoría de estos jóvenes, que trabajaron abnegada y desinteresadamente, eran parte de la generación de la Escuela de Ingeniería de la Universidad de Chile que tenían el lema "Poblete primero, Federici segundo y luego Pinochet", aludiendo a las autoridades impuestas en la Facultad, la Universidad y el país, las dos primeras depuestas por la presión y movilización de la comunidad universitaria.

El SRP fue probado varias veces antes del día "P" (5 de octubre) con tres marchas blancas e incluso con un apagón el día anterior que permitió verificar los equipos electrógenos. Los Centros de Acopio y Transmisión fueron cedidos por ONGs, empresas privadas y algunos particulares. Finalmente, el día "P" estuvieron resueltos todos los problemas. Solo un centro de acopio en Lota fue asediado y ocupado por fuerzas militares y los centros de Viña del Mar, Valparaíso y La Serena fueron visitados por autoridades militares.

EL CENTRO DE CÓMPUTOS

El computador central solo fue instalado dos días antes del 5 de octubre debido a que un local céntrico y cercano al Comando fue finalmente negado por los dueños. Se trató de un minicomputador Tower-32/600 con un procesador Intel 68.000, con un disco de 80 MB, con 18 terminales de entrada (dos de ellos como consolas de operación del sistema), una memoria RAM de 8 MB, una diskettera de 5 1/4" y un catridge para cintas y sistema operativo.

El software fue diseñado y programado por dos jóvenes ingenieros en computación en un par de meses de intenso trabajo. En un ambiente UNIX System V, se programaron en el lenguaje C aplicaciones multiprocesos con archivos de índices y datos compartidos utilizando acceso "single-writer/multiple readers". Los respaldos se realizaron en el INRIA (Francia) vía UUCP utilizando un modem Telebit Trailblazer de 9 Kbps [5]. El modelo de datos principal consideró dos tablas: de circunscripciones electorales (número, cantidad de mesas de mujeres y hombres, posiciones de comienzo en tabla de mesas) y de mesas (registro, circunscripción, número, sexo, y votos Sí, NO, nulos y blancos) [6].

EL DÍA "P"

El progreso del recuento de votos el mismo día 5 de octubre fue informado públicamente por el Ministerio del Interior, el Comando del NO y el Comité por las Elecciones Libres (CEL) que realizó un escrutinio rápido basado en un 10% de las mesas. La dinámica se puede observar en la tabla adjunta (**Figura 3**) con datos proporcionados por una delegación internacional de observadores presidida por Adolfo Suárez [7]. Puede notarse la prontitud del Ministerio del Interior por entregar primeros resultados favorables al SÍ (el primero con menos de un 1% de los votos) y su largo silencio posterior de cuatro horas.

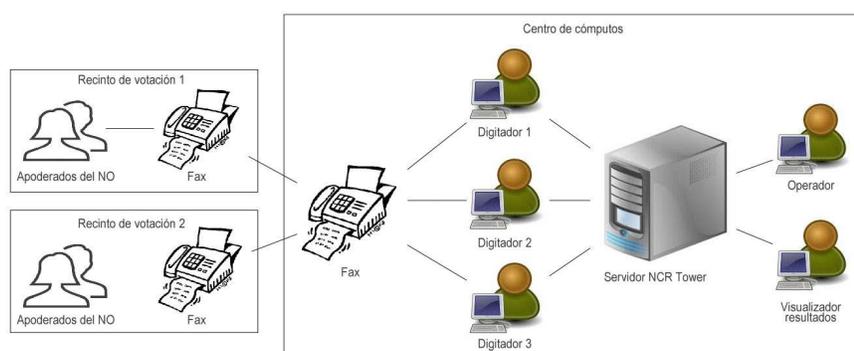


FIGURA 2
DISEÑO DE TRANSMISIÓN, PROCESAMIENTO Y DESPLIEGUE DE INFORMACIÓN.

Por otra parte, los resultados entregados por el Comando y por el CEL permitieron conocer oportunamente una buena aproximación del resultado final que días después entregó oficialmente el Tribunal Calificador de Elecciones (54.71% por el NO y 43.01% por el SÍ) y que mostró una variación de solo tres décimas con los resultados finales del Comando.

Cabe señalar que un equipo de jóvenes ingenieros de la Universidad Técnica Federico Santa María desarrolló un Sistema alternativo de Control Electoral para la “línea N” (adicional a la “línea O” descrita anteriormente). En este caso se utilizó una red de 85 PCs que utilizaron módems para transmitir la información a través de líneas telefónicas y entregaron a las 02:00 AM cerca del 75% de los resultados mesa a mesa [8].

Hora	Informante	% SÍ	% NO
19:10	Ministerio del Interior	57.36	40.54
21:30	Ministerio del Interior	51.3	46.5
21:45	CEL (735 mesas)	42.6	55.2
22:00	Ministerio del Interior	54.1	45.8
22:00	Comando	35.0	60.0
22:30	Comando	39.87	57.8
23:15	CEL (1600 mesas)	42.75	54.97
02:00	Ministerio del Interior	44.0	53.3
02:15	Comando	42.0	57.0

FIGURA 3

PUBLICACIÓN DE RESULTADOS EL DÍA 5 DE OCTUBRE DE 1988.

CONCLUSIONES

El Sistema de Recuento Paralelo del Plebiscito de 1988 fue una exitosa experiencia de construcción de un sistema socio-técnico. En efecto, además de la satisfactoria utilización de la tecnología de entonces por parte de un equipo de jóvenes profesionales, hay que destacar la participación de decenas de miles de personas también voluntarias: apoderados de mesas, apoderados generales, enlaces (“chasquis”) y técnicos. Todos ellos merecen un reconocimiento por su entrega y valentía y por protagonizar una verdadera hazaña en un trascendente hito para la recuperación de la democracia en Chile.

REFERENCIAS

- [1] Martner, Gonzalo. “El arquero que atajó”. En libro “La campaña del NO vista por sus creadores”. Editorial Melquíades. Agosto 1989.
- [2] Urquiza, Alberto. “El recuento paralelo”. En libro “La campaña del NO vista por sus creadores”. Editorial Melquíades. Agosto 1989.
- [3] Martner, Gonzalo; Urquiza, Alberto; Saavedra, Hernán; Saintard, Maurice. “Diseño y organización del Sistema de Recuento Paralelo del plebiscito de octubre de 1988”. Actas del XII Taller de Ingeniería de Sistemas. Departamento de Ingeniería Industrial, Universidad de Chile. Julio 1989.
- [4] Quintana, Germán; De Saint Pierre, Didier; Navarro, Juan Claudio; Lessegneur, Marcelo. “Red de transmisión y procesamiento de la información en el recuento paralelo del plebiscito de octubre 1988”.
- [5] San Martín, Marcelo. “Presentación en Conversatorio”. Octubre 2018.
- [6] Navarro, Juan Claudio. “Presentación en Conversatorio”. Octubre 2018.
- [7] “La transición chilena hacia la democracia: El plebiscito presidencial de 1988”. Informe de la delegación internacional. 1989.
- [8] Díaz, Eduardo; Suárez, Andrés; Díaz, Óscar; Díaz, Jorge; Guerrero, Carlos; Pizarro, Jorge; Haeger, José; Lefranc, Jean; Velez, Gastón. “CAOS: un sistema rápido de control electoral”. Actas del XII Taller de Ingeniería de Sistemas. Departamento de Ingeniería Industrial, Universidad de Chile. Julio 1989.



FIGURA 4
CONVERSATORIO 2018: GONZALO MARTNER, HERNÁN SAAVEDRA, DIDIER DE SAINT PIERRE, MARCELO SAN MARTÍN Y JUAN CLAUDIO NAVARRO.

CONMEMORACIÓN

El viernes 5 de octubre de 2018 se realizó un conversatorio en el DCC, donde participaron algunos de los principales protagonistas del desarrollo del Sistema del Recuento Paralelo: el Doctor en Economía, Gonzalo Martner; el Ingeniero Civil Eléctrico, Hernán Saavedra, y los ingenieros civiles en Computación Didier de Saint Pierre, Marcelo San Martín y Juan Claudio Navarro (Figura 4).

En la ocasión se exhibió una emotiva intervención de Germán Quintana, Ingeniero Civil Eléctrico y uno de los coordinadores del proyecto, quien no pudo asistir por su trabajo en el Banco Interamericano de Desarrollo (BID). Asistieron también varios de los entonces jóvenes ingenieros que participaron en el proyecto (ver Figura 5). Cabe señalar que el mismo día, en el Museo de la Memoria y los Derechos Humanos, se inauguró una sección dedicada al Plebiscito de 1988 incluyendo el Sistema de Recuento Paralelo. ■



FIGURA 5
DE PIE: CARLOS ENCALADA, GUILLERMO DÍAZ, RICARDO HERRERA, GONZALO MARTNER Y MARCO ZÚÑIGA. AGACHADOS: CARLOS ÁLVAREZ, HERNÁN SAAVEDRA, DIDIER DE SAINT PIERRE, JUAN CLAUDIO NAVARRO Y MARCELO SAN MARTÍN.



start-up

SILICON VALLEY

ENTREVISTA A

NICOLÁS LEHMANN

Por Federico Olmedo

INTROLOGIC



DCCC

INNOVACIÓN



Nicolás Lehmann es Ingeniero Civil en Computación y Magíster en Ciencias mención Computación de la Universidad de Chile. Apenas finalizó su magíster, se fue a trabajar a una empresa en Menlo Park, California, Estados Unidos. Nos contactamos con él para que nos cuente sobre su experiencia allí y sobre su paso por el DCC.

Para empezar Nico, cuéntanos con respecto a tu formación en el DCC. ¿Qué fue lo que te motivó a hacer el magíster? ¿Lo disfrutaste?

Siempre he tenido una inclinación hacia a la academia y el paso natural, finalizada la ingeniería, fue continuar con el magíster. Durante este período lo pasé muy bien y me gustó mucho el mundo de la academia: es entretenido estar pensando un problema por semanas, hasta que de repente se te ocurre “la idea”. Ese momento “Eureka” es muy satisfactorio.

Y terminado el magíster, ¿ya tenías pensado trabajar en Estados Unidos?

En realidad terminado el magíster no tenía idea qué hacer; tampoco había tenido mucho contacto con la industria, más que las prácticas profesionales. Pero sí tenía el deseo de salir del país, y de la nada me llega esta oferta de IntroLogic, en California. Como ya conocía el mundo de la academia dije, “bueno, ahora vamos a ver cómo es la industria” y así es que me vine a Estados Unidos.

Cuéntanos entonces un poco sobre IntroLogic, la empresa donde trabajas. ¿A qué se dedica? ¿Qué tamaño tiene?

Es una empresa relativamente pequeña. Somos siete trabajando acá y dos personas trabajando desde Chile. La empresa se fundó hace más o menos tres años, pero el producto actual lo empezamos a desarrollar a fines del año pasado, antes hacíamos algo totalmente distinto. Este producto que ofrecemos surge de la necesidad que tiene el departamento de ventas de toda compañía en identificar potenciales consumidores de sus productos. La manera tradicional de hacer eso es a través de lo que se llama *cold outreach*: le mandas mails a gente que no conoces con la esperanza de que se interesen en tu producto, y te respondan. Y la verdad es que a ningún vendedor le gusta enviar *spam*. Lo que nosotros ofrecemos hoy en día es completamente distinto; consiste en explotar las redes de contactos de los empleados de la empresa para identificar -entre ellos- los posibles consumidores. De esta manera estableces un vínculo a través de una presentación, y no a través de un correo o llamada telefónica *spam*. Esta solución la empezamos a idear hace un año y no se encuentra plenamente desarrollada, pero por el momento ha recibido una aceptación enorme, los vendedores se emocionan cuando le contamos la idea.

¿Y cómo se fundó la empresa? ¿Cuántas posibilidades de financiamiento hay en Estados Unidos para las start-ups?

La empresa fue fundada por un grupo de personas -mayormente de origen chileno- que habían estado trabajando en Groupon luego de la venta de una start-up anterior. Después de trabajar unos años ahí decidieron formar una nueva start-up y así nació IntroLogic, en ese tiempo con otro nombre y un producto totalmente diferente. Las facilidades para formar start-ups aquí son impresionantes: es muy fácil buscar inversionistas y hay mercado para todo. En Chile, cualquier idea que tengas, por muy buena que sea, el mercado es súper chico.

¿Y cómo crees que te formó el DCC para el trabajo que están haciendo ahora?

Creo que el DCC me formó súper bien. Como te comentaba antes, cuando estudiaba siempre tuve una inclinación académica y me dediqué a aprender algoritmos, lenguajes... y eso ayuda mucho cuando estás trabajando porque provee más herramientas analíticas para poder predecir qué tan bien se va a comportar algo. Pero aun así, uno se enfrenta a muchas cosas que en la universidad nunca ve. Construir software y que la gente lo use tiene sus complejidades... pero el DCC sin duda te prepara súper bien.

Pasando a lo extralaboral, ¿cómo es la vida en Silicon Valley? ¿Te costó mucho acostumbrarte?

La verdad es que me acostumbré súper rápido. Encuentro que Chile es bien parecido a Estados Unidos... quizás porque Chile se quiere parecer a Estados Unidos (risas). Cosas simples como las marcas más comunes que uno ve son las mismas acá que allá. La gente aquí es muy simpática y hay muchos latinos. Y acá uno hace las mismas cosas que hace en Chile: salir con amigos, ir a bares, salir a andar en bicicleta.

Por último, ¿qué le dirías a alguien que esté terminando la carrera y también esté evaluando irse a Estados Unidos?

Le recomendaría a ciegas que se venga para acá. Hay muchas posibilidades y el trabajo resultó mucho más entretenido de lo que uno puede encontrar en Chile. Lo fundamental es perder el miedo y venir para acá, que no es tan difícil en verdad. Uno siempre puede venir a probar, y si no te gusta, regresas. Por suerte, para nosotros que trabajamos en computación, las posibilidades de encontrar trabajo son siempre altas. ■

DOCTORADOS DEL DCC

FRANCISCO GUTIÉRREZ



Para nadie es sorpresa que la población mundial está envejeciendo a un ritmo acelerado. Dado esto, los servicios públicos, tanto en países desarrollados como en vías de desarrollo, están desesperadamente buscando formas para mitigar el potencial efecto negativo que trae consigo el cuidado de adultos mayores en una escala mayor a la que actualmente somos capaces de hacerlo. En particular, el rol de la familia surge como una estructura de apoyo y contención para hacer frente a este problema.

En la actualidad, el diseño de tecnología computacional para apoyar el cuidado de adultos mayores no ha reconocido ampliamente la complejidad y heterogeneidad de dicha actividad. Consecuentemente, esto introduce una brecha entre las necesidades de los adultos mayores y los servicios que sus redes de apoyo son capaces de proveer. Por ejemplo, la mayoría de los adultos mayores en Latinoamérica no han sido capaces -ni han manifestado interés- en adoptar tecnología digital para interactuar socialmente con sus familias. Esta actitud, junto a un alto grado de compromiso derivado

de la fuerte obligación filial característica de las sociedades colectivistas, sobrecarga a los familiares más implicados en el cuidado informal de sus adultos mayores, generando indirectamente tensiones con otros miembros. Luego, en lugar de tecnología más avanzada para apoyar el envejecimiento en el hogar, se requiere de un mejor entendimiento de la complejidad y diversidad de las experiencias de vida y necesidades de cuidado de los adultos mayores, para así apoyar el diseño de soluciones socio-técnicas potencialmente efectivas.

Mi tesis estudia empíricamente, a través de un enfoque inductivo basado en trabajo cooperativo asistido por computador, el escenario de interacción y colaboración en familias del Cono Sur. La finalidad de dichos estudios es modelar, de manera integral, la estructura y dinámica de los procesos de cuidado informal e interacción intergeneracional para comprender sus sutilezas desde un enfoque socio-técnico, y así diseñar nuevos mecanismos basados en computador para intervenirlos. Esto se traduce concretamente en identificar aspectos de diseño claves de dichos procesos, considerando el punto de vista de los adultos mayores y cómo se articulan los distintos miembros de la familia en este contexto. En otras palabras, el entender las implicancias de diseño derivadas del análisis sistemático del escenario de estudio, nos lleva a formular consideraciones de diseño que deben ser tenidas en cuenta a la hora de concebir soluciones



DE IZQUIERDA A DERECHA: PROFESORES JOSÉ A. PINO Y SERGIO OCHOA; FRANCISCO GUTIÉRREZ; PROFESORES JULITA VASSILEVA (EN LA PANTALLA - UNIVERSITY OF SASKATCHEWAN, CANADÁ), GONZALO NAVARRO, MARCELA RODRÍGUEZ (U. AUTÓNOMA DE BAJA CALIFORNIA, MÉXICO) Y BÁRBARA POBLETE.

que apoyen dichos procesos. Organizando estas consideraciones de diseño de acuerdo a su relación con las partes involucradas y las dimensiones que inciden en la arquitectura del software a desarrollar, nos lleva a proponer guías estructuradas para apoyar el proceso de diseño, desarrollo e implantación de sistemas computacionales en el dominio de estudio. Finalmente, mi tesis propone una serie de instrumentos que ayudan, por un lado, a identificar consideraciones y guías de diseño relevantes, y por otro, a apoyar a diseñadores de software a tomar decisiones arquitectónicas informadas a través de un proceso guiado en base a evidencia empírica.

Puede sonar cliché, pero realizar un doctorado no es tarea fácil. Hay que invertir una cantidad de tiempo no menor en leer, trabajar, escribir, volver a leer, volver a escribir, enjuague y repita... uno de los desafíos más importantes fue el tener que constantemente lidiar con lo imprevisto y digerir las críticas, desde las más constructivas hasta aquellas en donde te patean en el suelo (y, que en el fondo, de las que siempre se puede rescatar algo). Al respecto, no tengo sino

elogios para destacar la calidad humana, técnica y desafiante que se respira en el DCC. Quiero agradecer, en particular, al Profesor Sergio Ochoa, quien fue más que un mentor en formar mi espíritu crítico y desarrollo de una carrera científica en los más altos estándares. El DCC es, en mi opinión, un lugar ideal para hacer investigación de impacto y con sentido.

Después de finalizar mi doctorado, me he dedicado a consolidar mi carrera como investigador joven, continuando con la línea de trabajo que comencé en mi tesis y construyendo redes tanto a escala nacional como internacional.

En particular, uno de mis objetivos en el mediano plazo es promover el desarrollo de investigación de punta en Interacción Humano-Computador en Latinoamérica, una labor que estoy desarrollando junto al *Special Interest Group on Computer-Human Interaction (SIGCHI)* de la ACM.

DANIEL MORENO



Ingeniero Informático de profesión, llegué a Chile en 2010, justo después del terremoto. Hasta ese entonces, trabajé como desarrollador y jefe de proyectos en empresa, pero por esos giros del destino decidí perseguir un doctorado en el DCC, para el cual fui aceptado. Inicialmente no tenía un tema específico en mente, así que me dediqué a conversar con los profesores acerca de las áreas en que trabajaban. Es así como

conocí a Sergio Ochoa, quien eventualmente sería mi director de tesis.

En ese entonces, Sergio trabajaba en soporte de respuesta rápida ante desastres, entre otros temas. Con la cercanía del terremoto de 2010, el cual evidenció graves falencias en la infraestructura de soporte tecnológico móvil, el envío de mensajes localizados y el posicionamiento libre de infraestructura eran temas de alto interés. Así, el posicionamiento de dispositivos móviles se volvió mi área de investigación, con un pequeño twist: en vez de enfocarnos en mejorar una tecnología específica, construiríamos un modelo que determine qué estrategia de posicionamiento disponible es “mejor” para un dispositivo, dadas sus circunstancias tecnológicas.

Así nació CAMPOS (Context-Aware Model for Positioning on Outdoor Scenarios), un modelo de posicionamiento que “lee” el ambiente de un dispositivo y elabora una suerte de lista de estrategias que pueden ser usadas para realizar posicionamiento. La idea detrás de CAMPOS

es que cualquier dispositivo pueda determinar su posición independiente de sus características. Para ello, un dispositivo debe “saber” qué estrategias de posicionamiento puede usar de acuerdo a sus capacidades tecnológicas, y ser capaz de determinar si existe la infraestructura necesaria en el ambiente operacional para acceder a ellas.

Para construir el modelo usamos el ns-3, un entorno altamente versátil para la simulación de interacciones entre dispositivos móviles. Esta herramienta nos permitió diseñar y ejecutar distintos escenarios, cada uno con distintas condiciones ambientales y características de los participantes, entre otros. Los datos obtenidos fueron procesados y alimentados a un clasificador *random forest*, el cual determina el orden de prioridades de las estrategias de posicionamiento disponibles.

Un dispositivo que usa CAMPOS escanea su ambiente y detecta las características que puede usar para estimar su posición, las evalúa, y a partir de ellas escoge la estrategia de posicionamiento más adecuada. Si una estrategia recomendada no puede ser accedida por cualquier motivo, el modelo intentará con las siguientes estrategias, en orden de prioridad. De esta forma, un Smartphone con lector RFID y receptor GPS podría usar GPS como primera opción, RFID como opción secundaria, y en caso de que no pueda acceder a ninguna, intentará estimar su posición colaborativamente por triangulación, usando dispositivos cercanos con posiciones conocidas como puntos de referencia.

Y eso con la tesis. En cuanto a mi experiencia en el DCC, fue muy entretenida, aunque agobiante. Para mí el doctorado fue como ir a



la guerra; al principio no sabía lo que pasaba, solo hacía lo que tenía que hacer, hasta que de un momento a otro las cosas se complicaron rápidamente y entramos en terreno desconocido... ahí fue cuando empezó la verdadera experiencia académica, ese 0.001% que uno le aporta a la ciencia. Por suerte, tanto la planta docente y administrativa como los compañeros del doctorado estaban siempre dispuestos a ayudar (y en algunos casos, a “carretear”).

Y en cuanto a mí, siempre creí que terminando el doctorado volvería a trabajar en la empresa, pero como alguna vez me dijo Sergio: “*El doctorado te cambia la vida, socio...*” Ahora trabajo como profesor conferenciante en la Escuela de Ingeniería Civil en Computación de la Universidad de Talca, dedicado a docencia e investigación en temas relacionados con ingeniería de software.



DE IZQUIERDA A DERECHA: PROFESORES NELSON BALOIAN, LEANDRO NAVARRO (UNIVERSITAT POLITÈCNICA DE CATALUNYA) Y SERGIO OCHOA; DANIEL MORENO; PROFESORES NANCY HITSCHFELD Y GONZALO NAVARRO.

LUIS SILVESTRE



Soy Licenciado en Informática mención en Sistemas de Información de la Universidad Mayor de San Andrés (Bolivia), Magíster en Ciencias mención Computación y recientemente Doctor en Ciencias mención Computación del Departamento de Ciencias de la Computación (DCC) de la Universidad de Chile.

Mi tesis de doctorado se enmarca en el área de ingeniería de software dirigida por modelos, cuyo título es “A domain specific language to support the definition of transformation rules for software process tailoring”. Este trabajo de investigación doctoral estuvo guiado por los profesores del DCC, María Cecilia Bastarrica y Sergio Ochoa. La principal motivación de realizar dicha investigación se relaciona con mi trabajo

realizado en el magíster y el interés propio de realizar investigación aplicada en la industria de software. El primer acercamiento para realizar mi tesis de doctorado se produjo a partir de la participación e interacción con los profesores del DCC que dirigían el proyecto FONDEF – ADAPTE, cuyo principal objetivo fue desarrollar herramientas basadas en ingeniería dirigida por modelos para la adaptación automática de procesos de software a contextos específicos de proyectos.

La Ingeniería Impulsada por Modelos (MDE en inglés) es un nuevo paradigma de desarrollo de software que tiene la intención de mejorar la construcción de software elevando el nivel de abstracción mediante el uso de modelos y transformaciones. La adaptación de procesos de software es la actividad de adaptar el proceso de software de una organización a las necesidades de proyectos particulares. Recientemente, MDE se ha aplicado para adaptar procesos de software

utilizando modelos y transformaciones, pero su uso no es fácilmente aplicable en la industria del software. Esto porque el principal problema de la adaptación basada en MDE es la generación de las reglas y transformaciones, porque se requiere un conocimiento en lenguajes específicos de transformación como ATL o QVT. Como un medio para abordar el problema, mi trabajo de investigación consiste en equilibrar la formalidad requerida por la ingeniería impulsada por modelos y la usabilidad que necesitan los usuarios para aplicar MDE en la industria. Para ello, se propone una definición formal de reglas que permitan una generación sistemática y automática de transformaciones, y la construcción de herramientas que permitan a los usuarios aplicar la adaptación de procesos sin la necesidad de interactuar con modelos y transformaciones.

Mi trabajo de investigación presenta tres aportes. Primero un lenguaje de dominio específico (DSL en inglés), el cual permite definir formalmente reglas de transformación como modelo, segundo una transformación de alto nivel (HOT en inglés) que toma como entrada las reglas de transformación como modelo y genera como salida transformaciones en código ATL, y finalmente, se construyó un conjunto de herramientas integradas (ATAGeTT), que permiten a los usuarios generar modelos, reglas y transformaciones a partir de interfaces gráficas. ATAGeTT encapsula toda la complejidad de generar modelos, reglas y transformaciones para aplicar la adaptación de procesos de software de forma usable y útil a los potenciales usuarios sin la necesidad de interactuar con los formalismos de MDE.

Quiero agradecer al DCC, la Universidad de Chile, CONICYT y al Gobierno de Chile por permitirme realizar mis estudios de doctorado. Durante el doctorado en el DCC tuve la oportunidad de formarme como investigador, aprender lecciones de vida y conocer a gente maravillosa en Chile. Como investigador pude participar en varias conferencias y workshops importantes, e incluso en la ACM Student Research Competition; lo que me motiva a seguir con la línea de investigación de MDE+adopción en la industria e investigación empírica. Como lecciones de vida no tengo palabras para expresar mi gratitud a mis profesores guía María Cecilia Bastarrica y Sergio Ochoa por sus valores humanos, las conversaciones sobre la vida, las charlas sobre todas las líneas de investigación y la paciencia con mi forma especial de "hacer ciencia". Finalmente me es imposible no agradecer a Angélica Aguirre, Sandra Gáez, Francia Ormeño y Christian Bridevaux (Programa de Educación Continua del DCC), que hicieron de mi estancia en el DCC muy placentera y acogedora.

Desde marzo de 2017 soy académico de la Universidad de Talca, en Curicó. Mi plan académico involucra equilibrar investigación con docencia. Para ello pretendo postular a concursos de investigación de CONICYT, así como también a concursos regionales del Maule para continuar con la evolución de ATAGeTT. Además, pretendo continuar aportando con la formación de profesionales en computación desde la Región del Maule mediante la actualización de contenidos en el área de ingeniería de software del Departamento de Computación de la Universidad de Talca. ■



DE IZQUIERDA A DERECHA: PROFESORES SERGIO OCHOA, GONZALO NAVARRO Y BEATRIZ MARÍN (UNIVERSIDAD DIEGO PORTALES); LUIS SILVESTRE; PROFESORES MARÍA CECILIA BASTARRICA Y YADRAN ETEROVIC (PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE).

Postgrado y Educación Continua en el **DCC**



Diplomas de Postítulo en:

- Ciencia e Ingeniería de Datos
- Gestión de Calidad de Software
- Gestión de Proyectos Informáticos
- Ingeniería de Software
- Ingeniería y Calidad de Software
- Inteligencia Artificial
- Seguridad Computacional
- Tecnologías de la Información

☎ : + 56 22 9784965
+ 56 22 9784976
✉ : +56 9 6406 4274

pec@dcc.uchile.cl



Doctorado en Computación

Acreditado hasta septiembre de 2024

Magíster en Ciencias mención Computación

Acreditado hasta mayo de 2023

☎ : +56 22 9784891
+56 22 9784330

estudios@dcc.uchile.cl



Magíster en Tecnologías de la Información (vespertino)

Acreditado hasta diciembre de 2022

☎ : + 56 22 9784965
+ 56 22 9784976
✉ : + 56 9 6406 4274

pec@dcc.uchile.cl





Bits

DE CIENCIA

www.dcc.uchile.cl
revista@dcc.uchile.cl



fcfm

Ciencias de la
Computación
FACULTAD DE CIENCIAS
FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE