

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

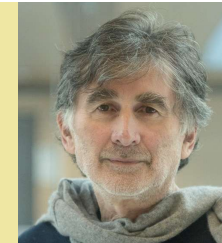
éxito en el último tiempo. Sin embargo, una desventaja de este método es su operación tipo “caja negra”, es decir, es complejo precisar las motivaciones detrás de cada inferencia hecha por un sistema de IA. Comprender el tipo de información y los procesos que usan estos modelos en su operación es un 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.

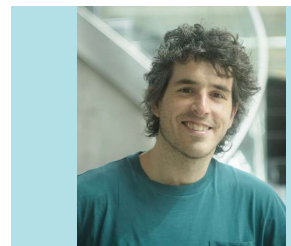
Investigadores asociados:



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.



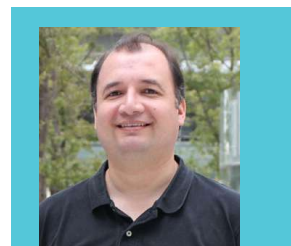
Jorge Pérez
 Doctor en Ciencias de la Ingeniería, Pontificia Universidad Católica de Chile; académico de la Universidad de Chile.



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



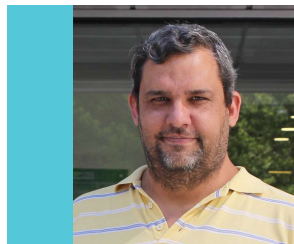
Éric Tanter
 Doctor en Ciencias de la Computación, Universidad de Chile y de la Universidad de Nantes, Francia; académico de la Universidad de Chile.



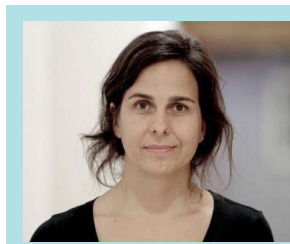
Benjamín Bustos
 Doctor en Ciencias Naturales, Universidad de Constanza, Alemania; académico de la Universidad de Chile.



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.



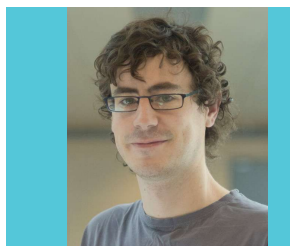
Marcelo Mendoza
 Doctor en Ciencias mención Computación, Universidad de Chile; académico de la Universidad Técnica Federico Santa María.



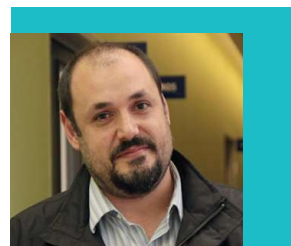
Sebastián Valenzuela
 Doctor en Comunicaciones, Universidad de Texas en Austin, Estados Unidos; académico de la Pontificia Universidad Católica de Chile.



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.