

Chile será sede del encuentro más importante de ciencia de datos a nivel mundial

- *Por primera vez en sus más de 50 años, el evento internacional líder en ciencia de datos se llevará a cabo en A. Latina, teniendo a Chile como país anfitrión. La cita, co-organizada localmente por la Pontificia Universidad Católica e Inria Chile, reúne a los expertos de la academia y la industria que están a la vanguardia en el estudio y manejo de datos digitales a gran escala.*

Santiago, Chile. Mayo, 2024.- En junio de 2024, las miradas de todos los especialistas en *data science* a nivel global estarán puestas en nuestro país. Durante seis días de trabajo intensivo **se llevará a cabo en Santiago la conferencia SIGMOD/PODS 2024**, en la que más de 800 científicos provenientes de universidades, centros de investigación y empresas de Chile y el extranjero abordarán –tanto a nivel teórico como aplicado– los desafíos en el manejo de grandes volúmenes de datos.

“No es por azar que se escogió a Chile como sede. Hace ya varias décadas que la comunidad chilena tiene una importante presencia en esta conferencia, con al menos 10 investigadoras e investigadores que viven en el país y que habitualmente presentan sus resultados científicos en SIGMOD/PODS. Varios de estos trabajos han recibido importantes premios en este encuentro”, señala **Pablo Barceló**, director del Instituto de Ingeniería Matemática y Computacional de la P. Universidad Católica de Chile (IMC UC), investigador del Instituto Milenio Fundamentos de los Datos (IMFD) y del Centro Nacional de Inteligencia Artificial (Cenia), y co-director del comité organizador local de SIGMOD/PODS 2024.

Nayat Sánchez, directora de Inria Chile y co-directora del comité organizador de SIGMOD/PODS 2024, enfatiza que “el esfuerzo de traer esta SIGMOD/PODS a Chile no ha sido menor y el que lo hayamos logrado es un reconocimiento al trabajo de científicos que son considerados como líderes en la región y el mundo. Ese prestigio ganado por años ha hecho que nuestro país sea sede de esta conferencia, un hito que nos permitirá impulsar aún más el posicionamiento de Chile en el área de datos”.

Barceló agrega que “es desde Chile de donde han surgido investigaciones que hoy permiten, por ejemplo, extraer información para entender el comportamiento en redes sociales en tan solo segundos”.

En SIGMOD/PODS también participan expertos de empresa tecnológicas que no solo siguen de cerca los avances el área de ciencia de datos, sino que desarrollan su propia investigación aplicada, como Amazon, Apple, Huawei, Microsoft, Google, Alibaba y Oracle, que –además– son sponsors oficiales de la conferencia.

CIENCIA HECHA EN CHILE EN SIGMOD/PODS 2024

De las **cuatro conferencias magistrales del evento, dos están a cargo de chilenos**. **Ricardo Baeza-Yates**, profesor titular de la Universidad de Chile e investigador sénior del Instituto Milenio Fundamentos de los Datos (IMFD), abordará los desafíos y limitaciones en el campos de los datos y el *machine learning*. En tanto, **Marcelo Arenas**, profesor titular de la P. Universidad Católica de Chile e investigador asociado del Instituto Milenio Fundamentos de los Datos (IMFD), hablará sobre su reciente trabajo en cómo proveer explicaciones a las decisiones tomadas por los modelos de inteligencia artificial.

También se presentarán los resultados de innovadoras aplicaciones y estudios hechos por científicos de nuestro país. Los **trabajos chilenos se enfocan en las bases de datos de grafos**, modelos que almacenan información interconectada entre sí, lo que permite que –al hacer consultas– los resultados sean más rápidos y precisos.

Una de las presentaciones mostrará **avances que buscan mejorar los sistemas de búsqueda sobre bases de datos de grafos**, para que arrojen no sólo información precisa, sino también más rica en profundidad y matices. Los investigadores son Diego Arroyuelo y Juan Reutter (P. Universidad Católica de Chile, IMFD); Benjamín Bustos, Aidan Hogan y Gonzalo Navarro (Universidad de Chile, IMFD), y Adrián Gómez-Brandón (Universidade da Coruña, España, IMFD).

También se presentará **MillenniumDB, un nuevo motor de búsquedas para bases de datos de grafos** que ya ha demostrado ser de dos a 10 veces más rápido que otros sistemas en uso actualmente (como los de Amazon o Neo4j). En el desarrollo de esta innovación participan 14 académicos, investigadores e ingenieros, y fue liderado por Domagoj Vrgoč y Carlos Rojas (P. Universidad Católica de Chile, IMFD).

Aidan Hogan (U. de Chile, IMFD) y Domagoj Vrgoč (P. Universidad Católica de Chile, IMFD) estarán a cargo de una **sesión tutorial** en donde **mostrarán cómo hacer consultas a bases de datos de grafos a gran escala**.

Otra innovación que se presentará es **REmatch**, una herramienta con la capacidad de **extraer información a partir de un patrón desde documentos de texto**. REmatch fue desarrollada por Cristian Riveros y Domagoj Vrgoč, junto con Vicente Calisto, Gustavo Toro y Nicolás Van Sint Jan, todos de la P. Universidad Católica de Chile, y Kyle Bossonney (Universidad de Oxford).

LOS NOBELES DE LA COMPUTACIÓN

En la larga historia de SIGMOD/PODS se han presentado avances que hoy resultan esenciales para que exista el comercio en línea, los motores de búsqueda, las redes sociales y la inteligencia artificial. Tan relevantes son las innovaciones presentadas en esta conferencia que, en sus cinco décadas, cuatro científicos han recibido el Turing Award, también conocido como el “Nobel de la Computación”.

En los 70, en los primeros años de la conferencia, los esfuerzos se centraron en llevar a la práctica los trabajos fundamentales realizados por Edgar F. Codd, creador de las bases de datos relacionales. Estos sistemas son utilizados por todas las industrias y sectores que manejan grandes volúmenes de información: la banca, los sistemas de compras en línea, los registros de salud, la gestión de inventarios en el retail, y muchísimos más. En el mundo de la ciencia de datos, Codd marcó un hito y su investigación le hizo merecedor en 1981 del Turing Award.

En la década de los 80, Jim Gray profundizó la investigación de Codd abordando el problema de la integridad de las bases de datos. Su trabajo fue clave para crear sistemas que permitían la consulta de información por parte de múltiples usuarios al mismo tiempo y está hoy presente en aplicaciones como el procesamiento de transacciones bancarias en línea y el comercio electrónico. Gray recibió el Turing Award en 1998.

El más reciente laureado es Michael Stonebraker, que desarrolló nuevos modelos y sistemas de manejo de bases de datos que revolucionaron el mercado, como Postgres, que permite conectar información más compleja y diversa. Sus innovaciones son aplicadas por empresas y servicios como Instagram, TripAdvisor, Uber y Spotify, entre muchos otros. Por su innovación en este terreno, Stonebraker recibió el Turing Award en 2015.

EL IMPACTO DE LA WEB Y LAS REDES SOCIALES

Las siguientes décadas fueron sumando más nombres ilustres: **en los 90, la aparición de la *world wide web*** generó nuevos campos de investigación, como –por ejemplo– la creación de sistemas para extraer e intercambiar información desde datos a gran escala, en donde

conflúan millones de usuarios. En este ámbito **destaca el trabajo del mexicano Héctor García-Molina, de la Universidad de Stanford (EE.UU.)** quien fue **mentor del proyecto** de dos estudiantes de doctorado que revolucionaron internet: **Sergei Brin y Larry Page, creadores del motor de búsqueda Google.**

“Los **2000s es la década donde se acuñó el término Big Data** y se identifican los desafíos que conlleva su manejo, como volumen de la información, velocidad de producción y diversidad de los datos. Esto **impulsa el trabajo en nuevos métodos y sistemas para campos como la astronomía o la secuenciación del ADN**”, explica Pablo Barceló.

Desde 2010 en adelante, la irrupción de poderosos algoritmos de inteligencia artificial (IA) que utilizan como base grandes repositorios de datos ha hecho que parte de la comunidad de científicos que se reúnen en torno a SIGMOD/PODS se esté enfocando en el gran desafío del sesgo en IA.

“Se ha hecho urgente la necesidad de desarrollar investigación en el uso de datos y en IA de manera responsable, que se haga cargo de los riesgos que tiene la manipulación de grandes volúmenes de información para tomar decisiones. Muchos de los trabajos presentados en SIGMOD/PODS abordan de manera teórica y aplicada estos desafíos”, concluye Nayat Sánchez.

Para contacto:

*Paula Silva, Encargada de Comunicaciones
Instituto Milenio Fundamentos de los Datos (IMFD)
E: paula.silva@imfd.cl / C: +56 9 5110 2915*

*Noemí Miranda, Directora de Comunicaciones
Instituto Milenio Fundamentos de los Datos (IMFD)
E: noemi.miranda@imfd.cl / C: +56 9 9879 9427*

*Costanza Cristino, Comunicaciones
Inria Chile (Agencia Parla)
E: coni.cristino@parla.cl / C: +56 9 9289 5205*