

De izquierda a derecha: Aidan Hogan, Claudio Gutiérrez, Pablo Barceló, Bárbara Poblete y Jorge Pérez.

NÚCLEO MILENIO

Centro de Investigación
de la Web Semántica (CIWS)

Por Pablo Barceló B. (Co-director) y Claudio Gutierrez G. (Investigador asociado)

¿Web o Internet?

Es común aún confundir internet con la web. El primero es un sistema de enlaces físicos (red de comunicación alámbrica e inalámbrica con un protocolo llamado TCP IP) a nivel planetario. El segundo es un espacio lógico de información, una suerte de diario mural universal donde todas las personas pueden publicar y leer. Es decir, internet se refiere a una infraestructura material de comunicación; la web a un espacio virtual de conocimiento.

La web, una poderosa combinación de innovación técnica y participación social, está transformando el mundo en que vivimos. Esto se nota particularmente al nivel de creación y difusión del conocimiento, de la comunicación y socialización de la información, y del procesamiento de datos. Fue concebida como un espacio de participantes autónomos y distribuidos, por oposición a los sistemas tradicionales de datos y conocimiento que son centralizados y acotados.

La web nace de la impredecible expansión (a una escala antes nunca vista para procesos sociales) de tres nociones muy simples: El protocolo de comunicación (HTTP), la noción de identificadores universales distribuidos (URLs, URIs), y un lenguaje de presentación de texto (HTML). La web hoy día se ha extendido al uso de multimedia (imágenes, video, etc.), diferentes tipos de formatos de datos (XML, RDF, CSV, entre otros) e interacción social.

Web como objeto científico

Es tentador pensar que el estudio de la web podría reducirse al de áreas científicas ya existentes, por ejemplo, bases de datos, sistemas distribuidos, redes y comunicaciones,

ingeniería de *software*. Sin embargo, al poco andar la comunidad internacional advirtió que este era un objeto con características únicas y muy complejas. En el caso de la web es un grafo dinámico, que se actualiza distribuidamente, y cuyas dimensiones y complejidad escapan a todo lo que haya antes conocido la humanidad. La investigación de este objeto impone la necesidad de modelos radicalmente distintos a los anteriores (teoría de redes complejas, lógicas de concurrencia masiva, procesamiento de *petabytes* de información, interfaces adaptables, automatización de la semántica, universalización de los sistemas de metadatos).

Progresivamente la web se ha ido transformando, desde sus orígenes como un repositorio universal de documentos enlazados, a un espacio de información y conocimiento. Podemos considerar la web actual como un gran espacio de bases de datos enlazadas por medio de relaciones semánticas. Es decir, sitios como Google Scholar, DBLP, arXiv.org y PubMed son fuentes de datos mutuamente enlazadas que conforman una red semántica de artículos científicos, información sobre sus autores y afiliaciones, relaciones de citación.

Web semántica

Para darle un marco de coherencia lógica y de solidez científica a lo anterior, es necesario desarrollar el soporte semántico para hacer entendible los recursos de la web. Usando una metáfora, es necesario crear un sistema de carreteras en este mundo virtual, estandarizarlas, dotarlas de simbología, de apoyos y metadatos para poder orientarse en ellas, así como entender que los objetos que se acceden por medio de ellas (hoteles, restaurantes, playas, museos, hospitales, escuelas) necesitan tener un significado. Debido a las dimensiones de la web, es

imprescindible que este proceso de dotar y aprovechar la semántica de la web se realice de manera semiautomática.


El proyecto antes descrito se conoce mundialmente como “Web Semántica”. Uno de los grandes desafíos en la actualidad es darle fundamentos lógicos y científicos a este proyecto. La investigación de nuestro núcleo –Centro de Investigación de la Web Semántica (CIWS)– se enmarca en este desafío. Nuestro estudio se enfoca en temas cruciales para el desarrollo del proyecto, como son la relación entre web semántica y bases de datos, el desarrollo de lenguajes lógicos que permitan consultar la semántica de la web, el estudio de modelos de datos de grafos, la búsqueda semántica, y el análisis en línea de la información publicada.

Nuestro centro

Actualmente Chile posee uno de los grupos más visibles a nivel mundial en las áreas de Fundamentos de la Web Semánticas y Fundamentos de Bases de Datos. El Núcleo Milenio Centro de Investigación de la Web Semántica no solo viene a consolidar el trabajo mancomunado de nuestro grupo en los últimos años, sino que nos permitirá seguir afrontando como actores principales los desafíos de la web de datos del futuro.

Cabe destacar que el grueso de los investigadores principales del Núcleo Milenio Centro de Investigación de la Web Semántica son académicos del Departamento de Ciencias de la Computación (DCC) de nuestra Facultad. Es el caso de Pablo Barceló, quien es uno de los investigadores responsables del núcleo, Claudio Gutiérrez y Jorge Pérez, quienes son investigadores asociados, y Aidan Hogan, Bárbara Poblete y Benjamín Bustos, que participan como investigadores jóvenes. Además, en torno al núcleo se reúnen varios postdocs y alumnos de postgrado del DCC. A través del Núcleo también establecemos contacto con investigadores de otras instituciones, en particular, la Pontificia Universidad Católica de Chile y la Universidad de Talca.

Entre los logros más importantes de nuestro grupo en los últimos años están los siguientes: (1) Establecimos el puente entre los conceptos de Web Semántica y Bases de Datos (C. Gutiérrez, C. Hurtado, A. Mendelzon, Jorge Pérez: *Foundations of Semantic Web databases*. JCSS 77(3), 2011) (2) Desarrollamos la formalización sistemática de un lenguaje para la publicación, extracción y procesamiento de datos semánticos a escala web (J. Pérez, M. Arenas, C. Gutiérrez. *Semantics and complexity of SPARQL*. ACM TODS 34(3), 2009). (3) Fuimos pioneros en establecer los fundamentos de un modelo de datos abstracto para la especificación de propiedades semánticas en grafos (P. Barceló. *Querying graph databases*. PODS 2013). (4) Construimos prototipos para sentar las bases de qué es un buscador semántico al nivel de la web (A. Hogan, A. Harth, J. Umbrich, S. Kinsella, A. Polleres, S. Decker. *Searching and browsing Linked Data with SWSE: The Semantic Web Search Engine*. Journal of Web Semantics 9(4), 2011). (5) Analizamos el nivel de credibilidad de la información publicada en la web (Carlos Castillo, Marcelo Mendoza, Bárbara Poblete: *Information credibility on twitter*. WWW 2011).

La visibilidad del trabajo realizado por los investigadores del Núcleo se ve reflejada de diferentes formas: Nuestros artículos suman más de 7000 citas en Google Scholar, hemos recibido premios a los mejores artículos en las conferencias más importantes del área (WWW, ISWC, ESWC, PODS, ICDT), permanentemente se nos invita a dar charlas plenarias en los congresos más destacados (PODS, ESWC), nuestros artículos se publican en las revistas más emblemáticas de Computación, Bases de Datos y Web Semántica. 

Enlace relacionado:
<http://ciws.cl/>