

# Entrevista Nacional:

## Ricardo Baeza - Yates

Por Gonzalo Navarro



Es, por lejos, el investigador chileno más citado en ciencia de la computación, y el único investigador del área que pertenece a la Academia de Ciencias de Chile. Fue el creador del Centro de Investigación de la Web (CIW), único Núcleo Milenio en la disciplina, y es responsable del establecimiento del primer laboratorio Yahoo! Research en Chile y Barcelona. Hoy es vicepresidente de Investigación para Europa y Latinoamérica de Yahoo! Research, y supervisa los dos laboratorios ya mencionados más uno reciente en Haifa, Israel. Además es un viajero incansable, de actividad frenética en general, sin que por ello deje algo de tiempo para una buena conversación. Quisimos conocer de primera mano la visión de este profesor del DCC sobre la Web, la ciencia de la computación y Chile en particular.

*Se ha dicho que investigar en la Web es algo único, en el sentido de que por un lado presenta desafíos algorítmicos, matemáticos y de sistemas formidables, y por otro puede tener un impacto inmediato en la sociedad. Como alguien que comenzó su carrera en algoritmos, ¿qué es lo que te atrajo, y qué es lo que te atrae ahora, de investigar en la Web? ¿Los temas de investigación han cambiado con el tiempo o son los mismos? ¿Sigue estando el foco en lo algorítmico o se ha movido a lo estadístico? ¿Cuáles son los desafíos más importantes en investigación en la Web para los próximos años?*

La Web es actualmente el conjunto de datos más grandes que ha producido la humanidad, pues en la práctica uno puede generar un número infinito de páginas a través de páginas dinámicas. En septiembre de 2008 ya había más de 180 millones de servidores

Web y posiblemente más de 20 mil millones de páginas estáticas. Su complejidad se debe a su volumen, continuo cambio y diversidad. Lo que me atrajo inicialmente fue la posibilidad de que las tecnologías de búsqueda que me interesaban pudieran ser usadas por muchas personas, por ejemplo a través del buscador de todo Chile, TodoCL. En otras palabras combinar la teoría y sus aplicaciones. Los temas relacionados con búsqueda de información siguen existiendo pero aparecen otros nuevos y uno de ellos es entender mejor la Web, donde la estadística y la minería de datos son importantes. Pero eso no significa que los temas anteriores dejen de ser importantes. Por eso los desafíos actuales son una mezcla de retos ya conocidos (recolectar la Web, jerarquizar páginas, etc.) y entender mejor la Web para poder aprovechar el conocimiento implícito que las personas generan, ya sea aportando

contenido (en particular vía Web 2.0) o usando la Web.

*Desde 2002 diriges el Núcleo Milenio Centro de Investigación de la Web, el único núcleo Milenio en computación en Chile, y en 2006 abriste el laboratorio de Yahoo! Research en Chile. ¿En qué crees que estas iniciativas han impactado o esperas que impacten en el país. ¿A nivel científico, social, educacional? ¿Ves diferencia entre el estado de la investigación en la Web antes de 2002 y ahora? ¿En qué sentido dirías que se ha avanzado?*

A nivel científico el CIW permitió crear el grupo más grande de investigación en computación en un tema específico, sobrepasando la masa crítica. Este grupo ha logrado impacto internacional y por supuesto nacional. Por ejemplo, cuando

Gary Marchionini, conocido investigador de la Universidad de Carolina del Norte, nos visitó hace algunos años, me comentó que él pensaba que teníamos uno de los mejores grupos del mundo. De hecho, la producción de artículos actualmente supera varias veces a la de 2002. Un hito importante para hacer investigación de punta en algunos temas fue el buscador de Chile, TodoCL, que comencé el año 2000,

*¿Qué fortalezas y debilidades tiene montar un laboratorio como el de Yahoo! Chile? ¿Cómo lo compararías con la situación en Barcelona o en EE.UU.? ¿Cuáles son los desafíos para la investigación en computación en Chile, qué hay que perseguir para estar al nivel del primer mundo? ¿La falta de masa crítica es un problema serio?*

*los buscadores, ¿se puede hacer mucho mejor que lo que haría un buen buscador horizontal? ¿De qué manera el estudio de una Web puede revelar información sobre la sociedad subyacente?*

Las Web regionales tienen algunas características que son locales, como el idioma y el contenido relacionado con la cultura local. Incluso en el tema de los buscadores se pueden hacer algunas cosas mejor como recorrer la Web en forma más completa y más rápido al estar más cerca y hacer mejor ranking de los resultados al ser la colección más homogénea.

Los estudios de la Web local, además de servir para ver su evolución, pueden ser usados para muchos fines, pues la Web es un reflejo de la sociedad y tiene información social y económica de un país. Con herramientas de extracción de información uno podría mejorar la Web (por ejemplo creando nuevas páginas en forma automática en base a contenido existente). También mediante minería de uso en un buscador regional uno podría recolectar el conocimiento implícito que existe en ese uso, lo que hoy se llama sabiduría de la gente. En particular aprovechar la Web 2.0 (anotaciones, opiniones, etc.) para extender la Web, que sería para mí el inicio de la Web 3.0.

### *¿Cómo defines la Web 3.0?*

Hoy hay diferentes definiciones según el autor, muchas de ellas disponibles en Wikipedia. En mi opinión, la Web 3.0 va a suponer el aprovechamiento del contenido y el uso que está generando la Web 2.0 para distintos objetivos, desde extracción de conocimiento hasta generación automática de contenidos. Todo esto será posible gracias a la minería web, que permite capturar la experiencia y el conocimiento de la gente y ponerlo al servicio de todos, consiguiendo al final una Web mayor y mejor.<sup>BITS</sup>



y permitió tener datos que pocos grupos de investigación académicos tenían. Todos estos hechos fueron fundamentales para convencer a Yahoo! de tener una sede de investigación en Chile, siendo el principal el capital humano.

A nivel social hemos tenido impacto a través de iniciativas que socializan la investigación como la Ventana Digital, los estudios de la Web Chilena y recientemente con el libro *Cómo Funciona la Web*. A nivel educacional hemos introducido temas nuevos que han permitido formar recursos humanos capacitados en temas diversos como tecnología de buscadores, algoritmos de compresión o Web semántica. A nivel gubernamental el laboratorio de Yahoo! se usa permanentemente como ejemplo para atraer otras iniciativas similares.

Chile está lejano del resto del mundo y es difícil traer investigadores de otros países. Tampoco es un país grande que pueda tener una fuente de talento local de gran tamaño. Este es el problema principal y la falta de masa crítica es siempre un problema importante. Los desafíos para hacer una buena investigación son similares a los de un país desarrollado, con el agravante que a veces la infraestructura y otros recursos no están disponibles a los niveles adecuados. Sin embargo estos obstáculos se pueden vencer con esfuerzo y tesón, algo que el DCC ha hecho a lo largo de más de 25 años.

*¿Existe alguna particularidad especial en una Web regional, como la chilena o la latinoamericana, que valga la pena intentar explotar o comprender? En el caso de*