

Necesidad de Semántica en la Web Chilena

Ernesto Krsulovic-Morales, Claudio Gutiérrez

{ekrsulov,cgutierr}@dcc.uchile.cl

Centro de Investigación de la Web,
Departamento de Ciencias de la Computación,
Universidad de Chile

Resumen

Presentamos la tendencia creciente de incorporación de semántica en la Web y discutimos las ventajas y perspectivas de su incorporación en el sistema de TI chileno.

Ponencia relevante a los temas de: Productividad en el sector público; Aplicaciones novedosas en Educación y Salud; TICs y democratización

1 Introducción

La *Web* es una red mundial de información que logra integrar los documentos que están contenidos en *Internet*. Esta integración se traduce en la accesibilidad universal a los documentos que ella contiene y se potencia mediante la utilización de hipertexto. Los hipertextos están soportados por estándares de servicio y codificación de textos, denominados *http* y *html* respectivamente. El nivel más primitivo de búsqueda en este mar de documentos consiste en integrar los documentos

accesibles en la Web en una gran base de datos documental. Esta idea, llevada a escalas antes impensadas, es lo que logran los motores de búsqueda hoy día. Los buscadores obtienen copias de los documentos en la Web y los “ordenan” o “catalogan” haciendo *rankings* de palabras claves que aparecen en los documentos de acuerdo a algoritmos de citas y criterios sintácticos.

El problema que algún día fué el difícil acceso a la documentación existente a nivel global, hoy se transformó en su contrario: la sobreabundancia de información. Y con ello apareció la necesidad de evaluar la calidad de la información. Un aspecto clave de este proceso es capturar el *significado (semántica)* de cada documento. La idea de introducir semántica a la Web apareció de la mano con el nacimiento de la misma. Tim Berners-Lee, el inventor de la Web, plantea en el año 1994, junto con la fundación del Consorcio de la Web, la necesidad de semántica en la Web [1].

Como frutos del trabajo realizado en los primeros años de funcionamiento en el W3C,

se propuso establecer una sintaxis común para especificar cualquier tipo de documento. Surge así XML. Más tarde, debido a la falta de un estándar, XML sería ampliamente aceptado como el lenguaje común para el intercambio de información. Junto a XML el W3C lanzó su campaña de transformar a la Web en una gran base de conocimientos, creando para ello un lenguaje de metadatos para estandarizar la descripción de significados, conocido como RDF. Similarmente a lo que ocurrió con XML, el lenguaje RDF ha sobrepasado los objetivos para los que fue creado y hoy día, el Consorcio de la Web lo considera un lenguaje para “proveer interoperabilidad entre aplicaciones, que intercambian información entendible por máquinas, utilizando la Web” [2]. Esto gracias a la flexibilidad y expresividad del modelo de datos que provee RDF. El proyecto de la *Web semántica* es ambicioso. El objetivo final es obtener agentes inteligentes, los que mezclados con técnicas de procesamiento de lenguaje natural, podrán hacer realidad el sueño de convertir el espacio universal de información que hoy constituye la Web en una gigantesca base de conocimientos distribuida.

Las motivaciones para incorporar semántica a la Web son múltiples. Probablemente todas tienen como base el distribuir tanto la información como el conocimiento y democratizar el acceso a ellos. Esto ya lo conocemos en el ámbito de la ciencia en la Web, con las bases de datos abiertas, la posibilidad de compartir información en línea hoy es real, como los proyectos SCED, GEON para intercambio de información de movimientos telúricos y

geológicos, vocabularios para anotaciones médicas como MeSH, los proyectos como *Semantic Grid*, etc. También el área del comercio notó hace tiempo las ventajas de un tal enfoque, donde hoy se cuentan con iniciativas de estandarización como ebXML, y a nivel semántico iniciativas como el vocabulario para intercambio de modelos de procesos de negocios BPML, la estandarización de vocabularios para recursos humanos XML-HR, o los vocabularios para administrar empresas del DMTF. En el área de Gobierno electrónico es porobablemente donde más se note la importancia de lograr completa interoperatividad a nivel de documentos. Más adelante mostraremos el caso chileno en detalle. Finalmente, en el área propiamente social, estas iniciativas recién están mostrando todo su potencial. Un buen ejemplo son iniciativas como FOAF, la sindicación y el blogging.

Estas iniciativas no sólo pueden ayudar a reducir costos a través de compartir servicios y datos, sino que sobretodo sientan la base para compartir información y colaborar, dos actividades en la base de cualquier proceso de sinergia que conduce a ideas y soluciones nuevas.

2 El caso de la Web Chilena

Desde un punto de vista global, la Web chilena es una componente sin gran importancia de la Web universal. Sin embargo, su importancia estratégica a nivel nacional en ámbitos

como la organización y democratización del Estado, difusión y participación cultural, comercio y relaciones industriales, por nombrar algunas, sobrepasa cualquier estimación que hoy podamos hacer. La Web chilena es la red que hoy día permite conectar casi instantáneamente organismos del Estado, ciudadanos con el Estado, ciudadanos entre sí, organizaciones sociales, comerciales, educacionales, etc. Buenos ejemplos son *SII*, *memorianacional.cl*, *chilecompras.cl*, *tramitefacil.cl*, servicios bancarios varios.

Sin embargo, a pesar de esta importancia, la organización de nuestra Web es pobre. No existe hoy día una política dirigida a “organizar” este gran espacio de información nacional. Podemos hacer una analogía con la red caminera nacional: una política dirigida permite no sólo lograr interconectar diferentes puntos geográficos, sino un sistema estandarizado de signos y señales permite su navegación. Hoy existen alguna políticas para desarrollar nuestra red Internet. Pero no hay ninguna para establecer un sistema mínimo de signos y señales que orienten la navegación, que den *significado* a cada uno de sus nodos. La prueba está en que *.cl* está particularmente mal catalogado. Buscadores globales como Google rankean mal nuestras páginas. Buscadores locales como TodoCl son insuficientes para los requerimientos que vayan más allá de una búsqueda sintáctica.

Desde el Estado recién comienzan los esfuerzos de estandarización en la dirección que discutimos. Esperamos que pronto XML será recomendación como estándar para documentación electrónica para los organismos del Estado. Hagamos notar que XML no se

menciona en el Libro Azul ni en informes recientes sobre el tema. Por otro lado, comparado con países como Estados Unidos o Finlandia, nuestra presencia en la Web es baja. Mientras antes incorporemos el tema de la semántica a la Web, podremos esperar mejores resultados. Recordemos que Tim Berners-Lee se lamentaba de no haber incorporado desde un comienzo en la Web alguna forma de semántica. Lo anterior significa esencialmente trabajo al nivel de modelamiento conceptual, de modelos de negocios, de modelos de intercambio de información entre los organismos del estado, de modelos de comunicación con los ciudadanos de parte del Estado, de modelos de comunicación de estos ciudadanos entre si, etc. En los hechos esto se traduce en creación de modelos conceptuales, creación de ontologías (vocabularios estandarizados), diseño de esquemas de documentación, que puedan ser reutilizados tanto para modelar procesos como también para guiar el desarrollo de software. Hagamos notar que estos modelos no sólo mejoraran la integración local, sino que aumentarán la visibilidad de nuestra información desde el resto del mundo.

El problema de obtener un acceso uniforme desde fuentes de datos disímiles heterogéneas es conocido como integración. La integración existe en diferentes niveles: desde la integración física que, como sabemos, es provista pasando por la integración sintáctica, hasta llegar a la denominada *integración semántica* que permite la integración automática de esquemas de información. Este es el nivel que se necesita para transformar la Web en una base de conocimientos [3]. Para ilustrar las ideas

anteriores, y mostrar la posibilidad real de implementarlas, mencionaremos brevemente algunas experiencias en el área de la educación, área que por motivos obvios es un muy buen caso de prueba para la introducción de ideas avanzadas. En Chile, hay algunos ejemplos, como el de Reuna con el proyecto SIEL ha completado un proceso de estandarización de metadatos para la descripción de recursos Universitarios. CONICYT se propone lo mismo para la comunidad científica, con el incipiente proyecto SICTI. *Cybertesis*, un proyecto de SISIB para la publicación en línea de memoria y tesis de las universidades chilenas. Entre nuestros proyectos en desarrollo [4], está el desarrollo una plataforma para la integración de recursos educacionales a nivel nacional. A nivel mundial existen numerosas actividades de estandarización de metadatos, todas orientadas a *e-learning*. Nuestro enfoque pasa por lograr describir las organizaciones, centrándonos en la reutilización de términos desde estándares reconocidos, para así lograr la integración de recursos educacionales que tienen presencia en *la red*.

[4] Grupo Metadatos Universidad de Chile, Septiembre 2003, <http://www.metadatos.cl>

Referencias

[Nota] En esta lista no aparecen referencias a los acrónimos utilizados en este artículo, los cuales están disponibles utilizando **Google**.

[1] *Plenary at WWW Geneva 94*, Septiembre de 1994. <http://www.w3.org/Talks/WWW94Tim/>

[2] *RDF Faq*, Agosto 2003.
<http://www.w3.org/RDF/FAQ#What>

[3] *Semantic Integration, Strategies and Tools*, Julio 2003, <http://www.topquadrant.com/documents/TQ0303.Semantic%20Integration.PDF>