

B.2. La web semántica es sólo una promesa

Por José-Ramón Pérez-Agüera

Pérez-Agüera, José-Ramón. "La web semántica es sólo una promesa".

En: *Anuario ThinkEPI*, 2008, pp. 34-38.



ES DIFÍCIL IMAGINAR un navegador semantizado cuando la inmensa mayoría de la Web, de los usuarios y de las aplicaciones aún no están semantizados.

La Web semántica aún parece una promesa de futuro, y corre el riesgo de pasarle como a muchos jugadores de fútbol, que se quedan en promesas y luego nunca llegan a nada. Existen los lenguajes, que de eso se encarga el W3C y compañía, pero no existe la información marcada o anotada con esos lenguajes, ni existe el software destinado a interpretar esos lenguajes. La mayoría del software que se usa hoy en día en Web semántica son prototipos destinados a la investigación, y aún están lejos, muy lejos, de poder ser usados en entornos industriales.

Si se prueba *Jena* o *Protégé*, dos viejos conocidos de todo aquel que trabaje en estas cosas, se podrán desarrollar entornos "de juguete" donde demostrar las excelencias de una Web Semántica, ¿pero qué ocurre cuando se intenta cargar una ontología con 100.000 instancias en *Protégé*? Si se hace la

La mayoría del software que se usa hoy en día en Web semántica son prototipos destinados a la investigación, y aún están lejos, muy lejos, de poder ser usados en entornos industriales

Se está empezando la casa por el tejado y eso puede dejar a la Web semántica en la larga lista de vaporware que ha poblado la web en los últimos 10 años

prueba, aquello empieza a no funcionar bien y va tan lento que dará la sensación de estar manejando un PC de los años 80. Algo muy similar pasa con *Jena*, y es que tanto uno como otro son, a día de hoy, incapaces de manejar las ontologías y los volúmenes de información necesarios como para ser funcionales más allá de un entorno de pruebas. Sólo solucionar los problemas de persistencia de este tipo de estructuras llevará varios años de desarrollo.

<http://jena.sourceforge.net/>

<http://protege.stanford.edu/>

Por otro lado están los razonadores, que son fundamentales para poder realizar inferencias sobre las ontologías, el ejemplo más conocido es *Racer*. Con este software pasa un poco lo mismo, hay pocos y encima los buenos suelen ser de pago, con lo cual su uso se complica aún más.

<http://www.sts.tu-harburg.de/~r.f.moeller/racer/>

Conclusión, a día de hoy no existe una infraestructura técnica real y robusta para la Web Semántica, ¿para qué sirven los navegadores semánticos si no tienen por dónde navegar? No es que no sea una idea estupenda, pero parece que se está empezando la casa por el tejado y eso puede dejar a la Web semántica en la larga lista de *vaporware* que ha poblado la Web en los últimos 10 años.

Terminaremos poniendo ladrillos

Por Jorge Serrano-Cobos



Comentaba **José-Ramón Pérez-Agüera** que se estaba haciendo la casa por el tejado, y que faltaban los ladrillos. Por lo que cuenta, los ladrillos serían la información marcada (los metadatos ¿estructurados? en ¿un? formato comúnmente utilizado) por la enorme mayoría

de usuarios/productores de información del mundo, y la argamasa un software con un razonador que pueda lidiar con el volumen resultante de información para “entender” la semántica generada entre la red de metadatos.

Bien, en cuanto a éstos, todos los días se ven florecer y apagarse nuevos formatos (desde los de *RSS* a los de *Mpeg-7*), y ya se sabe que hoy por hoy la norma entre los metadatos es que “cuanto más sencillo, más aceptación”.

Por otro lado observamos progresivamente más intentos, velados o abiertos, de obtener de los usuarios algunos de esos datos, véase *Google maps* ahora personalizado, que puede capturar la información que personalizemos de ubicaciones favoritas, *Google base*, en su día el ya extinto *Google catalogs* y ahora evolucionado, *Myspace* o *Facebook*, máquinas de recoger información personal, etc.

La personalización está por fin empezando a verse cada vez más, y eso es algo que les interesa a los grandes: quitarse *spam* (anunciantes enfocados a la masa), aumentar el *clickthrough* de sus anuncios (con anuncios más cercanos a tus intereses), y hacerlo de forma escalable (una vez más el rollo 2.0, dejando que seamos todos los que les demos los ladrillos).

Hoy por hoy la norma entre los metadatos es que “cuanto más sencillo, más aceptación”

Cuando lo que se obtenga de dar cierta información sea igual o supere el costo de darla, habrá (hay) mucha gente dispuesta a contarte lo que haga falta

Para recoger esa información lo que necesitan es darte una motivación: ésa es la clave. Cuando lo que se obtenga de dar cierta información sea igual o supere el costo de darla, habrá (hay) mucha gente dispuesta a contarte lo que haga falta, igualito que en televisión se ponen a parir maridos, suegras, etc., en público. Lo he comprobado con más de 300.000 usuarios reales. Lo mismo que el *ego-tagging*.

Por poner un ejemplo plausible, de no tardar tanto alguno de estos servicios podría permitirnos conocer dentistas entre tantos y tantos euros, o de tal aseguradora cerca de tu casa, que den el servicio exacto que necesitas, en el horario que te interesa, y que han sido bien valorados por otros usuarios (ontologías en la Web semántica o cosas como 11870). De eso a inferencias hay un trecho, ya, pero es un avance.

Me resulta todo esto similar al caso de la sindicación. Recuerdo que, en 2003, cuando empecé a hablar de ella pregunté abiertamente por qué no podía obtener algo más específico que el canal de información de todo un portal, y mientras se me decía que para qué, que si me agregaba el canal de la web que me interesaba ya era más que suficiente. Al poco ya teníamos aplicacioncillas que empezaban a permitir agregar los resultados de búsquedas específicas para estar actualizados constantemente (*Trucos de Google*); programadores unipersonales lo hicieron, incluso antes de que el propio *Google* pudiera generar sus alertas.

Quiero decir que, una vez más, después de la lucha entre dos tecnologías (*RSS* y *RDF*) han pasado años y, por un lado, ya es mucho más fácil agregar a los lectores de noticias búsquedas específicas (ejemplo) o sobre *tags*

El problema parece ser la sencillez de uso, tanto para aportar ladrillos como para usar la argamasa

concretos en *Technorati*; y, por otro, los lectores ya se entienden con diversos formatos y ya han sido incorporados en las nuevas versiones de los navegadores (*Firefox*, *Explorer 7*).

Así que, quién sabe si aquí podremos observar lo mismo. El problema es la sencillez de uso, tanto para aportar ladrillos como para usar la argamasa.

La velocidad de procesado, seguro que se va mejorando, y cuando se tienen miles de servidores...

En cuanto al papel del bibliotecario, pues sé que en su día **Tim Berners-Lee** jugó con la idea de usar la clasificación de la *Library of Congress* para la Web semántica (*automatic*

classification research) aunque el problema que le veo a usar cualquier clasificación de partida es que no será mundial (se dará de tortas con la *CDU*, con la rusa, con...) y a cada una le pasará lo mismo con sus sesgos y carencias.

Probablemente acabaremos colocando ladrillos. Ya lo hacemos. Sin intención de provocar *spam*, eso sí, como ha ocurrido con los *Dublin Core*.

Capacidad de cálculo, rapidez, sencillez, universalidad, escalabilidad, personalización, evolución... Veo mucha evolución. Como dije hace años por ahí en las redes: para mí lo interesante no es tanto el resultado final esperado como las mil ideas y posibilidades que pueden generarse en el camino. Como las ramificaciones de LA ontología, que quizá al final sea una ontología de ontologías, hechas con la mía, la tuya, la suya... Anda, ¡una ontología 2.0!

Ah, no, que esto era la web 3.0.

La Web | Imágenes | Maps | Noticias | Vídeo | Gmail | Más ▼

thinkipi@gmail.com | Mi perfil | Ubicaciones guardadas | Ayuda | Mi cuenta | Salir

Google Maps España

p. ej. "avenida de galicia, oviedo" o "bar de tapas en barcelona"
(biblioteca OR bibliotecas) Coruña

Buscar en el mapa | Buscar negocios | Cómo llegar

Resultados 1-10 de aproximadamente 552 para biblioteca or bibliotecas cerca de A Coruña -
Modificar búsqueda

Categorías: Bibliotecas, Fundaciones

- A Ayuntamiento De La Coruña** - más datos >
C/ Río Monelos S/N, 15006 A Coruña
981 184 298
Categoría: Bibliotecas
- B Biblioteca Municipal Infantil Y Juvenil** - más datos >
C/ Durán Loriga 10, 15003 A Coruña
981 184 388
- C Biblioteca Municipal Monte Alto** - más datos >
Vereda Polvorín 4, 15002 A Coruña
981 184 382
- D Biblioteca Sagrada Familia** - más datos >
C/ Antonio Pereira 1, 15007 A Coruña
981 184 392
- E Biblioteca Publica Do Estado** - más datos >
C/ Miguel González Garcés 1, 15008 A Coruña
981 170 219
- F Biblioteca Municipal De Castrillon**

Ubicación de las bibliotecas de La Coruña con Google Maps.

Normalización y experiencia analítica

Por **Belén Castelló**



Creo que la Web semántica pasa necesariamente por la normalización en la arquitectura de los contenidos de las páginas web. Sin una normalización en su elaboración no hay máquina por “inteligente” que sea que pueda “entender” lo que se le pre-

gunta y emular el pensamiento humano para dar una respuesta certera. Aun así, dudo que el factor humano sea jamás sustituido; pero no se trata de eso, sino de gestionar la incommensurable cantidad de información (y desinformación). Como decía, normalización; pero eso no es nuevo.

En cuanto al papel de los bibliotecarios: puesto que somos los profesionales que llevamos siglos clasificando contenidos, tenemos una tendencia (inducida por la práctica, supongo) a agrupar las ideas, los pensamientos en categorías, en clasificaciones, detectando fácilmente las conexiones internas que hay entre las manifestaciones aprehensibles del conocimiento humano.

Nuestra labor, a pesar de la invisibilidad de nuestra profesión (hace un par de años asistí en la *Facultad de Informática de la Universidad de Deusto* a un seminario sobre Web semántica y, siendo del área sólo otra bibliotecaria y yo, no se hizo en ningún momento alusión a los bibliotecarios y documentalistas ni a su trabajo en el mundo de las ontologías, clasificaciones de contenidos, o clasificaciones de contenidos de libros) es aportar nuestra experiencia analítica, de deducción y sobre todo de inducción, nuestra visión panorámica dentro de la empresa, institución, universidad o población en donde trabajamos porque de tanto clasificar contenidos, de tanto buscar respuestas tenemos una perspectiva privilegiada de cómo funciona el conocimiento.

Nuestra labor es aportar nuestra experiencia analítica, de deducción y sobre todo de inducción, nuestra visión panorámica dentro del entorno de trabajo

Podemos hacer mapas de conocimiento de nuestras comunidades de usuarios y de su ámbito, es decir, descendiendo al ejemplo concreto, pensando en voz alta: puedo hacer un mapa del conocimiento de la Administración en la que trabajo y una clasificación de su contenido, esto supone considerar cómo se organiza, qué necesita, qué ofrece, cómo le llamamos a todo ello en lenguaje natural y cómo le debemos llamar en un lenguaje controlado (¿habrá ontologías hechas a la medida de la Administración pública?, ¿habrá que hacerlas?). Ya se sabe lo que pasa con las sinonimias y homonimias, además de las distintas estructuras mentales de quien hace una página web y de quien busca en ella.

Como decía, ese lenguaje controlado para describir una organización y su “vida cotidiana” debería casar con otros similares, debería reflejarse en los metadatos de nuestra página web, y debería encajar con la “estructura mental” de los buscadores que nos van a encontrar, indizar y ofrecer como respuesta ante la búsqueda de una persona que en cualquier sitio del planeta pregunte sobre nosotros.

Este ejemplo puede ser extrapolable a otros ámbitos, incluso a proyectos a gran escala. Para resumir, debemos ofrecer nuestros servicios de clasificadores de contenidos y unido al trabajo de lingüistas, informáticos, psicólogos y sociólogos construir los puentes para que emisores y receptores de información a través de la Web se entiendan.

Belén Castelló
bibliobi-san@ej-gv.es

La contribución de los documentalistas

Por **Jesús Bustamante**



Coincido con **José Ramón** en que es necesaria una abundancia de contenidos marcados para que los navegadores semánticos cojan fuerza. También es necesario que uno o más estándares maduren y se extiendan lo suficiente como para que los pro-

ductores de contenidos los utilicen sin darle más vueltas. Es un poco una pescadilla que se muerde la cola.

En lo que se refiere al papel del documentalista, aquí van unas ideas sobre en qué puede contribuir:

- marcando, indizando contenidos;
- creando, mejorando y manteniendo vocabularios controlados (ontologías, etc.);
- interconectando vocabularios existentes (interoperabilidad);
- ayudando a diseñar sistemas que a quienes generen contenidos les permitan crear, asignar sus propios metadatos fácilmente.

Por ejemplo, podríamos contribuir a mejorar la usabilidad de las nubes de etiquetas tanto para su navegación como para la asignación de éstas a los contenidos. Una cosa que está ya clara es que los usuarios quieren contribuir y dejar huellas. Quieren por ejemplo marcar los contenidos. Las folksonomías por lo menos prueban eso.

Un par de ideas quijotescas para terminar (con punto seguido):

- Imaginad, por ejemplo, que pudiésemos también navegar por esas nubes de etiquetas siguiendo relaciones semánticas entre los términos.
- Imaginad que cuando incorporásemos un contenido a la sede web, éste recomendase términos (provenientes de vocabularios controlados existentes) basándose en el análisis semántico del texto incorporado, en el análisis de las etiquetas libres que ya hemos comenzado a asignar, y/o en términos asignados a contenidos que el sistema considere semejantes (por la razón que sea).

Yo creo que los documentalistas tienen mucho que aportar. Como siempre formando parte de equipos multidisciplinares (bien formados, motivados, bien pagados, y bien alimentados, por supuesto)

Jesús Bustamante

heyzeus@heyzeus.com

Recursos

Al hilo de este debate en ciernes..., y para aquellos que, como yo, tenemos una vaga idea de lo que es la Web semántica, viene muy a cuento este artículo que sirve como introducción para poder seguir esta interesante discusión con más conocimiento de causa.

Telefónica. "La Web Semántica, la siguiente generación de Webs". En: *Boletín de la sociedad de la información: tecnología e innovación*, 2007, abril, n. 94.

<http://sociedaddelainformacion.telefonica.es/jsp/articulos/detalle.jsp?elem=4299>

Y por cierto, también en este mismo *Boletín*, pero en un número anterior hay otro artículo titulado "Cómo será la web 3.0".

<http://sociedaddelainformacion.telefonica.es/jsp/articulos/detalle.jsp?elem=4215>

Lara Rey

lararey@ferroatlantica.es

Post sólo apto para "documentalistas híbridos" pero también para quien esté interesado.

Leído en diagonal parece un somero estado de la cuestión (por lo menos habla de los supuestos *must*), especialmente por la retahíla de comentarios:

"Web 3.0: When Web Sites Become Web Services"

http://www.readwriteweb.com/archives/web_30_when_web_sites_become_web_services.php

Mina Nabona-Jassans

Mina.Nabona@uab.es