
RECURSOS DE INTERNET

RECURSOS SOBRE LA WEB SEMÁNTICA

Lluís Codina*, Cristòfol Rovira*

Es ya un tópico afirmar que la mejor fuente de información sobre Internet es Internet. En el caso de la Web Semántica se cumple con creces. Aunque ya existe una abundante bibliografía impresa en forma tanto de monografías como de publicaciones periódicas, realmente, es imposible encontrar más información y *más autorizada* sobre el tema en otro sitio que en la propia Web. Una simple consulta en Google por el término en inglés (*semantic web*) arroja más de 33 millones de documentos.

Aunque es imposible determinar la exactitud de la respuesta (es sabido que, por ejemplo, no se puede ir más allá de la página 90 de resultados en Google), la cifra no deja de dar una idea aproximada de la cantidad de información sobre el tema presente en la Web.

La misma consulta, en la base de datos del ISI (a través de ISI Web of Science) que contiene básicamente referencias (y a veces el texto completo) de artículos de publicaciones impresas arroja un total de 1.120 resultados (consultas realizadas en febrero de 2006).

La cuestión es que, tal como destacábamos más arriba, resulta que algunas de las fuentes más autorizadas son, precisamente, las que pueden hallarse en la web, como es el caso de buena parte de los recursos que presentaremos a continuación.

En cualquier caso, dada la imposibilidad de presentar la guía “definitiva” sobre la Web Semántica, el lector puede interpretar este trabajo como una propuesta provisional que, posiblemente merezca ser complementada en futuras entregas.

Semantic Web Activity

<http://www.w3.org/2001/sw/>

Guía Breve de la Web Semántica

<http://www.w3c.es/divulgacion/guiasbreves/WebSemantica>

El proyecto de la Web Semántica ha sido impulsado por el World Wide Web Consortium (W3C), dentro de su dominio de actividades denominado *Technology & Society*. Puede considerarse, además, un proyecto personal del inventor de la Web y fundador y director del W3C, Tim Berners Lee.

* Universidad Pompeu Fabra, Barcelona. Correo-e: lcodina@gmail.com; cristofol@upf.edu.

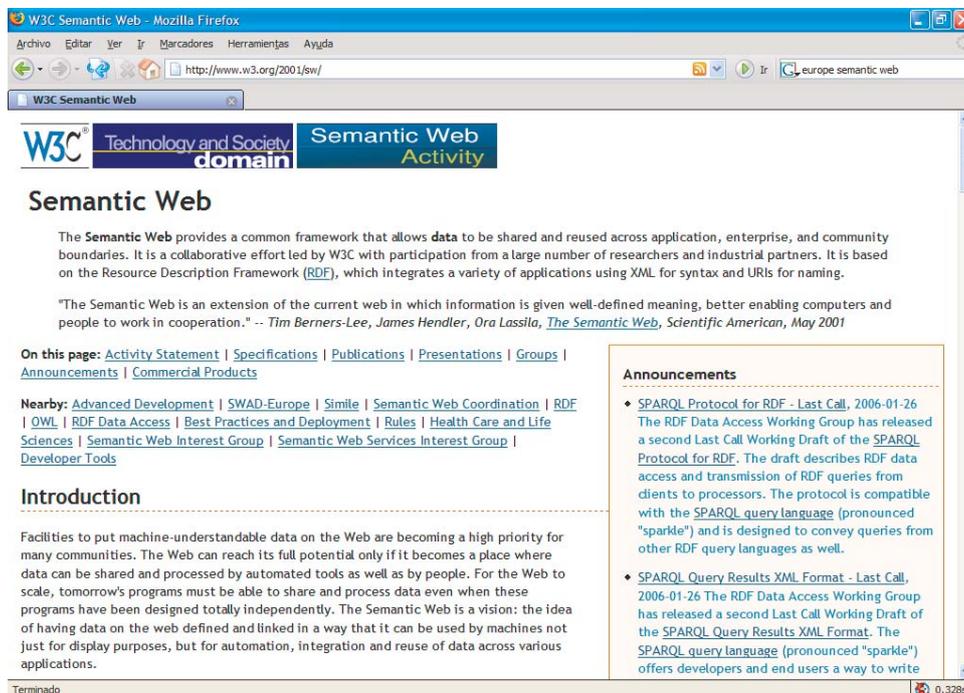
Expresado en forma muy compacta (y por tanto, haciendo saltar por el aire muchos matices) el proyecto pretende favorecer una nueva Web donde los contenidos posean valor semántico para los ordenadores conectados a la Red. De este modo, la Web Semántica es el paraguas unificador de un amplio conjunto de estándares, normas y especificaciones que trabajan en diversos niveles de la Web y de forma complementaria entre ellos.

El recurso que comentamos es el sitio oficial del proyecto, por tanto, no solamente es una de las mejores fuentes sobre el tema, sino en este caso, la fuente más autorizada y de conocimiento obligatorio para quien desee profundizar, ya sea por motivos profesionales o académicos en este importante campo de desarrollo futuro de la Web.

Se trata de un auténtico portal sobre la Web Semántica, ya que, desde aquí se puede acceder no solamente a las especificaciones y recomendaciones oficiales, sino también a artículos e informes, sitios relacionados y sitios de empresas que ofrecen tecnologías relacionadas con el proyecto.

Por supuesto, destaca el apartado referido a especificaciones (<http://www.w3.org/2001/sw/#spec>) ya que son el corazón de la Web Semántica. Por su parte, en la sede oficial para España del W3C (<http://www.w3c.es/>) pueden encontrarse algunos documentos traducidos al castellano como el indicado a la cabeza de este apartado.

Figura 1
Sitio web del proyecto de la Web Semántica del World Wide Web Consortium



Wikipedia - Web Semántica

<http://es.wikipedia.org/> > Web Semántica

El artículo de la Wikipedia sobre la Web Semántica es otro buen punto de partida, ya que después de ofrecer una definición bastante asequible y comprensiva propone una pequeña compilación de recursos sobre la misma que incluye enlaces a las diversas especificaciones relacionadas y una bibliografía muy escueta y básica, pero que incluye las obras fundamentales sobre el tema.

Semantic Web.org

<http://www.semanticweb.org>

Semantic Web.org es un portal sobre los aspectos de la web semántica más vinculados con el mundo de la ingeniería del software. Está soportado por un instituto de investigación de Irlanda consagrado a la Web Semántica (DERI) que tiene sede también en Austria y una delegación en la universidad de Stanford.

Es útil como lugar desde el cual es posible conocer convocatorias sobre jornadas y congresos, así como presenta también recopilaciones de herramientas, software, vocabularios y desarrollos sobre la web semántica.

XML

<http://www.w3.org/XML/>

XML.com

<http://www.xml.com>

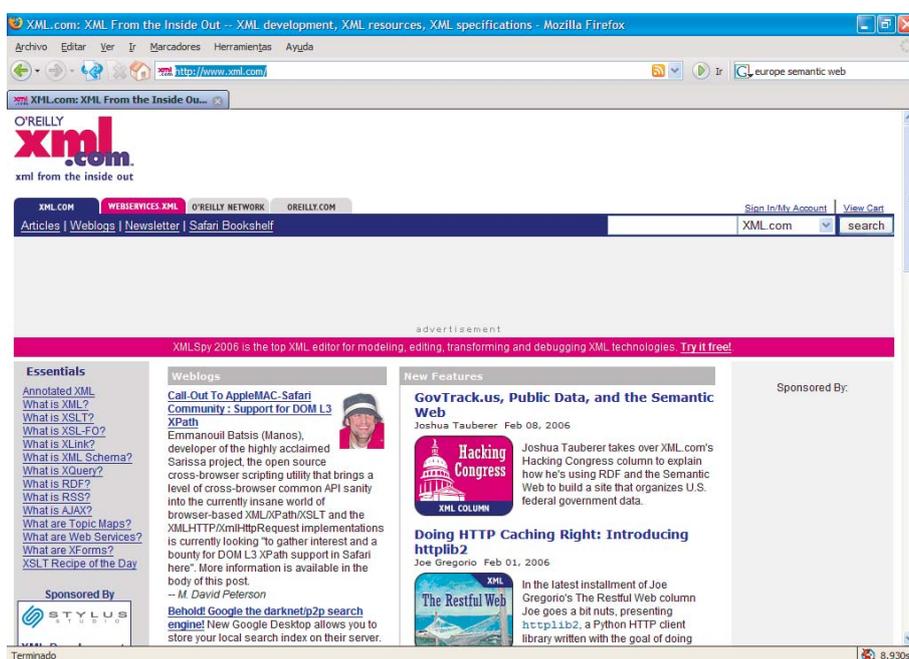
Es difícil a estas alturas pretender presentar XML, pero tampoco podían faltar en esta guía al menos dos de los principales recursos sobre el tema. Por tanto, aunque la presentación pueda pecar de ociosa, permítasenos recordar que XML es el lenguaje de creación de los lenguajes (un metalenguaje, por tanto) actuales de la Web. Hace tiempo que HTML ha dado paso a XHTML, es decir, un HTML codificado de acuerdo con las prescripciones generales de XML.

La cuestión es que XML es un conjunto de especificaciones de nivel abstracto que permite definir lenguajes para aplicaciones concretas, por ejemplo, publicar páginas web (XHTML). Con XML pueden generarse vocabularios especializados a cualquier nivel temático o de especificidad. Por ejemplo, existen aplicaciones XML para intercambiar noticias de actualidad, para expresar fórmulas matemáticas, para describir recursos, etc. XML permite generar y especificar los componentes de lenguajes de codificación de la información que transforman simples páginas de texto (o de texto más imagen) en auténticas bases de datos con campos bien delimitados y diccionarios de datos que precisan el alcance semántico, si es necesario, de cada línea de texto.

Otro de los aspectos clave de XML es que permite delimitar datos con tanta precisión y nivel de control que actúa como un formato universal de intercambio de datos entre aplicaciones y entre plataformas distintas.

Con un indudable sentido de la oportunidad, la editorial O'Reilly mantiene este portal sobre XML. Aunque con una finalidad indudablemente comercial, proporciona información abundante y de calidad sobre el tema.

Figura 2
Portal sobre XML de la editorial O'Reilly



Resource Description Framework (RDF)

<http://www.w3.org/RDF/>

¿Tiene sentido tratar aquí y por separado las diversas especificaciones y normas que componen la Web Semántica? Probablemente, no. Sin embargo, creemos que, al menos una de ellas sí lo requiere. Nos referimos a la especificación sobre *Resource Description Framework* (RDF).

RDF es uno de los pilares principales, junto con el lenguaje XML, de la Web Semántica. RDF es un lenguaje general de representación de información. Por decirlo de otro modo, es lo que, en nuestra terminología más reciente, denominaríamos un lenguaje de metadatos.

Los documentalistas estamos familiarizados con el concepto de metadatos desde hace muchos años (¿décadas, siglos?) bajo otras denominaciones anteriores: catálogos, documentos secundarios, clasificaciones, etc.

RDF es la respuesta de la ingeniería informática a la necesidad de disponer de sistemas altamente homogéneos para describir recursos digitales en particular y recursos de cualquier tipo en general. RDF no es un conjunto de metadatos predefinidos, como Dublin Core, sino un conjunto de especificaciones para crear cualquier tipo de metadatos siguiendo un conjunto de normas preestablecidas que hacen posible el intercambio de descripciones creadas por agentes distintos y sin necesidad de que se hayan puesto previamente de acuerdo entre ellos.

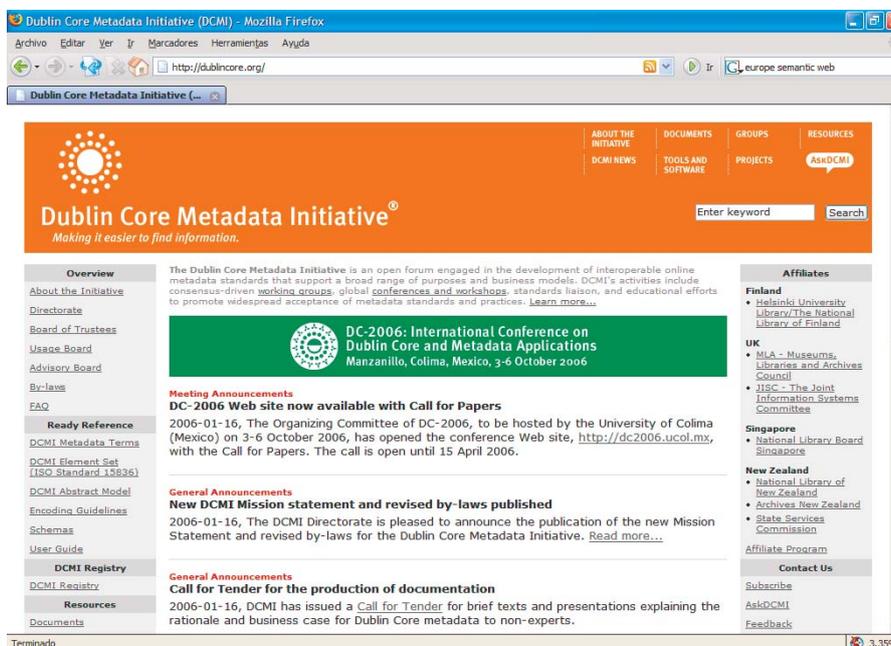
Dublin Core Metadata Initiative

<http://dublincore.org/>

La iniciativa de metadatos Dublín Core es anterior al proyecto de la Web Semántica, pero ha sido adoptada por parte del W3C y por diversos colectivos con intereses en el proyecto como uno de los conjuntos de metadatos predefinidos que pueden ser utilizados para describir recursos, en este caso, recursos digitales en línea, esto es, páginas y sitios web de cualquier naturaleza existentes en la Web.

En el sitio web de Dublin Core encontramos, por tanto, una valiosa información para conocer con detalle esta norma de metadatos y sus diferentes formas de codificación: XML, (X)HTML y RDF. Además, entre otros apartados incluye una sección de herramientas de software en línea o que pueden descargarse en el propio ordenador aplicar y usar Dublin Core en las páginas web de la empresa o corporación interesada en ello.

Figura 3
El sitio web de la iniciativa de Metadatos Dublin Core



Scientific American.com – The Semantic Web

<http://www.sciam.com> > May 2001 issue > The Semantic Web

En el sitio web de *Scientific American* se encuentra disponible en línea (y en acceso abierto) el famoso artículo de Tim Bernes Lee (escrito con Ora Lassila) del año 2001 presentando las ideas esenciales de la Web Semántica al gran público.

Para mucha gente fue el primer contacto con el proyecto y con el concepto de la Web Semántica y tuvo un gran impacto en el periodismo de divulgación científico de los dos años siguientes y en el imaginario social de quienes se movían en el ámbito de lo que podríamos llamar la cibercultura.

Tiene la virtud de elegir un enfoque tipo «gran público» (aunque tampoco sea para escolares precisamente), de modo que es una excelente introducción a la Web Semántica a «1.000 metros de altura».

Tiene el inconveniente (al menos para quienes firman esto) de constituir una presentación poco realista, ya que da por sentado el triunfo cercano de la Inteligencia Artificial (aunque no se la invoca explícitamente en el artículo). En todo caso, tiene el valor indiscutible de ser uno de los documentos «fundacionales» del proyecto y de tener como autor a un posible premio Nobel (la candidatura al Nobel de Bernes Lee está siendo promovida últimamente).

SWAD-Europe

<http://www.w3.org/2001/sw/Europe/>

Esta es la sede web oficial de un proyecto de la Unión Europea que tuvo su desarrollo entre los años 2002 y 2004. Aunque el proyecto ya no está en vigor, puede afirmarse que los informes generados durante el mismo conservan su interés al cien por cien.

Entre otros recursos disponibles destaca, probablemente, la colección de *reports*, entre los cuales destaca, a su vez, por su interés para el mundo de la Documentación, la colección de informes dedicados a RDF y lenguajes documentales como los tesauros (<http://www.w3.org/2001/sw/Europe/reports/thes/>).

RDF Thesaurus

<http://www.w3.org/2001/sw/Europe/reports/thes/rdfthes.html>

SKOS Core Guide

<http://www.w3.org/TR/2005/WD-swbp-skos-core-guide-20051102/>

Simple Knowledge Organisation System (SKOS)

<http://www.w3.org/2004/02/skos/>

Uno de los más afortunados resultados del proyecto europeo SWAD-Europe fue, entre otras cosas, el impulso de la aplicación RDF denominada Simple Knowledge

Organisation System (SKOS), cuya aceptación en la comunidad internacional de bibliotecarios-documentalistas no deja de aumentar gracias a que forma parte también de las especificaciones y recomendaciones del W3C.

Skos Core es un vocabulario RDF destinado a poder representar en formato RDF tesauros (o lenguajes documentales similares, como taxonomías). El resultado de codificar un tesoro en Skos Core es un archivo RDF. Esto permite que en lo sucesivo, la colección de tesauros existente se vaya codificando en el formato nuclear de la Web Semántica y, por tanto, significa la plena entrada de los tesauros en la Web Semántica.

OntoWeb

<http://www.ontoweb.org/>

Web Ontology Language Overview

<http://www.w3.org/TR/owl-features/>

Otro proyecto de la Unión Europea sobre la Web Semántica que tuvo su final en el año 2004 pero que ha dado resultados duraderos y de enorme calidad es OntoWeb, por lo cual hemos seleccionado esta web como una de las referencias importantes sobre el tema.

Las ontologías constituyen el aspecto más avanzado (¿más fantástico?) de la Web Semántica. Una ontología, en el sentido de la Web Semántica, es una colección coherente de aserciones sobre algún aspecto de la realidad. Estas aserciones incluyen declaraciones del tipo <un camión es un tipo de automóvil> y del tipo <un automóvil es un medio de transporte>; pero también restricciones como la siguiente: <un camión no sirve para transportar pasajeros>, o como la siguiente: <un camión no sirve para el transporte marítimo>, etc.

Dentro del proyecto de la Web Semántica existe un lenguaje de ontologías declarado canónico y que se denomina OWL. El sitio OntoWeb contiene abundante información sobre las ontologías así como análisis e informes sobre programas informáticos que permiten definir, crear y mantener ontologías. Afortunadamente, OntoWeb ha tenido continuidad con otro proyecto de la Unión Europea (Knowledge Web) que también dispone de su propio sitio web (<http://knowledgeweb.semanticweb.org/index.html>).

Protégé

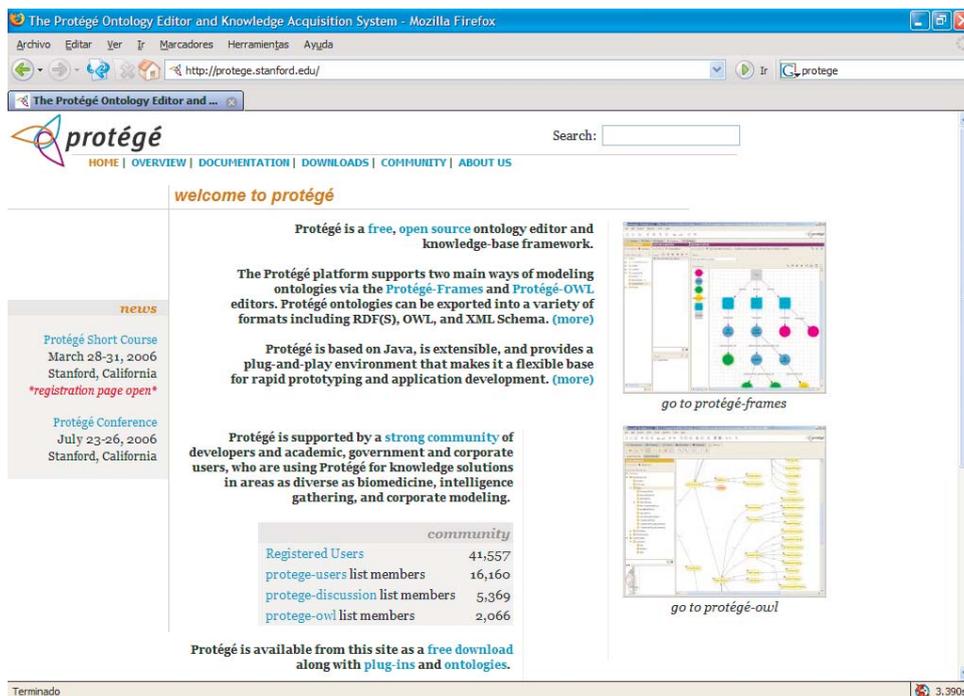
<http://protege.stanford.edu/>

Por muy incompleta que sea una guía de recursos sobre la Web Semántica, sería difícilmente justificable no incluir el sitio de la Universidad de Stanford dedicado a Protégé.

Protégé es un programa informático de tipo *open source*. Su función consiste en la creación, edición y mantenimiento de ontologías. Además de ser un programa exquisitamente diseñado en cuanto a interfaz y facilidad de funcionamiento, posee la capacidad de generar de manera automática el código fuente de cualquier ontología creada con Protégé en formato RDF/OWL.

En la web de Protégé podemos encontrar no solamente la manera de descargar el programa, sino diversos tutoriales sobre el uso de Protégé en particular y la concepción de ontologías en general. La mayoría de estos documentos tiene un aire claramente didáctico, como no podía ser de otro modo, tratándose de materiales de tipo didáctico.

Figura 4
Página principal del software open source Protégé para la creación y edición de ontologías



Como conclusión, podemos señalar que hemos privilegiado en esta breve guía los aspectos generales de la Web Semántica y aquellos que tienen mayor incidencia en el mundo de la documentación. Nos hemos limitado expresamente, además, a sitios web de carácter corporativo y no hemos pretendido incluir informes o artículos individuales, con la única excepción del documento «fundacional» de Tim Bernes Lee.

Posibles actualizaciones futuras de esta temática podrían centrarse en aspectos sectoriales, por ejemplo, las ontologías y recuperación de información, o aspectos verticales, por ejemplo, la Web Semántica aplicada a las bibliotecas o a la distribución de noticias, etc. Esperamos que, por el momento, y como una primera aproximación, los sitios reseñados puedan ser de utilidad a nuestro colectivo.