

[Página principal](#)[Nuestra profesión](#)[Socios patrocinadores](#)[Socios institucionales](#)[Cursos](#)[Publicaciones](#)[Grupos de trabajo](#)[Servicios a empresas](#)[Bolsa de Trabajo](#)[Relaciones externas](#)[Enlaces](#)[Mapa del web](#)**PUBLICACIONES**[Publicaciones](#) > [Boletín CLIP](#) > [Con Firma](#)> [Archivo de CLIP anteriores](#)

nº 41

2004

Con firma

**La Web Semántica:
una web "más bibliotecaria"**

Eva Mª Méndez Rodríguez

Dpto. Biblioteconomía y

Documentación

Universidad Carlos III de Madrid

emendez@bib.uc3m.es**Con Firma**

Eva Mª Méndez Rodríguez

*La Web Semántica: una web
"más bibliotecaria"***Nombres Propios**

Traslados

Nuevos socios

SEDIC AbiertoUn semestre de SEDICform@:
primer balance.....
SEDIC rumbo a la formación
integral de los profesionales.....
El proyecto europeo de
certificación de profesionales
en Información y
Documentación CERTIDOC
alcanza su recta final.....
Seminario Bibliotecas
accesibles en la Web:
Un reto urgente**Universidades**La ANECA publica el Libro
Blanco del Grado en

Información y Documentación

Centros en MarchaLa Universidad San Pablo-CEU
presenta la Biblioteca del Siglo
XXI**Créditos**

ISSN: 1137-0904

sedic@sedic.es

Al igual que ocurre en otras disciplinas, en la nuestra, la Documentación, cada cierto tiempo aparecen nuevos temas que, por su utilidad para nuestra profesión, por el potencial de desarrollo que tienen e incluso por su crípticismo designativo o por su complejidad conceptual, se ponen de moda. Esto es justamente lo que ha ocurrido con la Web Semántica (SW) que, a pesar de que no es un concepto tan nuevo, sino que responde, más bien, a la idea original que tenía Tim Berners Lee en 1989 de la Web, en los dos últimos años se ha convertido en un tema recurrente no sólo para informáticos y tecnólogos, sino también para los profesionales de la información que buscamos respuestas para responder al complejo mundo de la Web.

Explicar qué es la Web Semántica, máxime desde la carga editorial o de opinión que tiene este espacio que me ha otorgado SEDIC, no es una tarea fácil, sobre todo si tenemos en cuenta una serie de factores que dificultan aún más la comprensión de esta idea:

- > La propia expresión ya que, a pesar de ser afortunada en su traducción del inglés (*Semantic Web*), la yuxtaposición de los significados de los términos que la conforman (Web + semántica), no alcanza para explicar el intrincado concepto vinculado a esta expresión, ni mucho menos a explicitar o proyectar su verdadera importancia.
- > Hablar de Web Semántica implica hablar de múltiples *estándares* --o especificaciones técnicas con vocación de serlo-- representados por un conjunto interminable de siglas o acrónimos (tales como [RDF](#), [XML](#), [DCMI](#), [DALM](#), [OIL](#), [OWL](#), [SKOS](#), [RSS](#), [FOAF](#), [SVG](#), [CCPP](#), etc.) y de conceptos vinculados a ellos (metadatos, ontologías, agregación de contenidos, interoperabilidad, entre otros).
- > El aumento exponencial de la "sopa de letras" vinculada al desarrollo de la Web Semántica. Cada día aparecen nuevos *schemas* o *vocabularios* (que incluso se están empezando a llamar de forma genérica "ontologías"). Todos los días surgen nuevos esquemas, nuevas especificaciones a nivel disciplinar, que tratan de concretar la percepción de la SW, a través de vocabularios específicos RDF, una dimensión vertical, especializada y aplicable a un tipo de información y/o necesidad en la Web. (Se pueden consultar algunos de estos múltiples esquemas en: <http://www.schemaweb.info/schema/BrowseSchema.aspx>)
- > La falta de visibilidad de las tecnologías implicadas en la Web semántica, o mejor dicho, la excesiva transparencia que tiene para el usuario el uso de estas tecnologías. La SW está en desarrollo, pero son múltiples los proyectos y sistemas de información que utilizan ya nuevas o nuevas

proyectos y sistemas de información que utilizan, en mayor o menor medida, desarrollos en RDF: *bloggers*, redes sociales, agregadores y sindicadores de contenidos, e incluso buscadores como Kartoo <<http://www.kartoo.com>>.

- > El predominio del *middleware* en el desarrollo de aplicaciones, esto es, que la mayoría de las herramientas desarrolladas en este sentido, conllevan un nivel de software intermedio entre la infraestructura básica de la Web y las aplicaciones. Son muy pocas, por no decir insuficientes los software ejecutables para "crear" o "consumir" WS directamente; así existen distintas APIs (*Application Program Interfaces*), analizadores, validadores, etc. (Algunas de estas herramientas de la Web semántica están recogidas en: <http://www.w3.org/RDF/#developers>).

Si a todos estos factores le añadimos además: el entusiasmo y la proyección que sus principales instigadores --Tim Berners-Lee y James Hendler-- le han dado a esta idea, y la propia intangibilidad del concepto (que unas veces se acerca al cognitivismo y otras, a las más avanzadas teorías informáticas de Inteligencia Artificial, y siempre, a la disciplina que, de forma genérica, en Documentación, denominamos Organización y Representación del Conocimiento) la concreción y aprehensión de la Web Semántica, verdaderamente se complica. Pero... **¿qué es en realidad la Web Semántica? ¿por qué la Web Semántica es una web "más bibliotecaria"? ¿Estamos ante una realidad de evolución de la WWW o simplemente ante una moda o un "glamour" de carácter informático?**

Siguiendo a Berners-Lee, Hendler y Lassila en el que es ya el artículo electrónico sobre la [Web Semántica](#) más citado en este contexto, publicado en la revista *Scientific American* en mayo de 2001, la Web Semántica no es una nueva Web, es una extensión de la Web actual en la que la información se presenta con un significado bien definido, permitiendo a los ordenadores y a las personas trabajar conjuntamente. La creación de la Web Semántica o de las tecnologías y estándares que la posibilitan/arán, es una respuesta a la sensación que todos los que aplicamos la informática a la gestión de la información hemos tenido alguna vez, y que Daniel Connelly en su [The XML Revolution](#) verbaliza perfectamente en estos términos: *The bane of my existence is doing things that I know the computer could do for me*. A pesar de esta definición de SW y lo categórico de la reflexión de Connelly la Web Semántica es "muchas cosas" para "muchacha gente", pero siempre, un conjunto de tecnologías para la organización y representación del conocimiento digital que añaden semántica interpretable por las máquinas. Los documentalistas y bibliotecarios (o una buena parte de ellos) temen que la Web les exija más destrezas informáticas, conocimiento de lenguajes formales, XML/RDF, etc., mientras los informáticos persiguen una Web más bibliotecaria, con descripciones y sistemas de organización del conocimiento, que les permita crear una Web de datos con significado.

La Web semántica persigue una Web más inteligente, pretende, ni más ni menos, convertir la información en conocimiento, algo en lo que llevamos trabajando los documentalistas durante siglos. Para ello, se basa fundamentalmente en el marcado semántico y descriptivo no sólo de los documentos, sino también de los datos, a través de metadatos, información estructurada y legible automáticamente, sobre la información distribuida en la WWW, que proporcionan a los ordenadores una mayor capacidad para gestionar y recuperar esos datos. Hasta ahora, los motores de búsqueda, independientemente de que basen su funcionamiento en algoritmos refinados como el *page-rank* de Google, han trabajado, debido sobre todo a la parquedad semántica del HTML, con significantes (con términos), mientras que la potencialidad de esta "segunda generación de la Web", fundada sobre la flexibilidad de XML y la potencialidad de encontrar nuevos medios de expresión semántica de RDF, trabajará con significados, con conceptos. El componente básico de esa búsqueda por conceptos lo constituyen las ontologías que son algo así como la representación o especificación informática de una conceptualización; al fin y al cabo, metadatos orientados al contenido o a la recuperación por materias, algo que también llevamos haciendo los bibliotecarios mucho tiempo. Estas ontologías y/o tesauros requieren simplemente (o complejamente, según se mire), estándares dirigidos a la creación de vocabularios interpretados por la máquina (OWL o SKOS-Core, si hablamos expresamente de tesauros), más allá de los estándares dirigidos a la creación de vocabularios interpretables por el hombre (ISO 2788 ó Z39.19).

Aunque son múltiples las tecnologías implicadas en la arquitectura y

visualización de la Web semántica y no es, en absoluto, el objetivo de este artículo explicar cada una de ellas, ni siquiera, definir su alcance, sí que debo dedicar una pequeña reflexión al *Resource Description Framework (RDF)*, un estándar de facto en la construcción de la Web semántica, que supone una estructura contenedora de distintos modelos de metadatos (tanto descriptivos como orientados al contenido). RDF no es más que esa estructura que permite construir lenguajes lógicos que puedan funcionar juntos en la Web semántica, o bien codificar vocabularios ya existentes como el Dublin Core (DCMI-RDF) o dicho de otra forma, una manera de usar XML orientado a los datos (y a los metadatos) y no sólo a los documentos.

A pesar de esta aparente complejidad, los principios básicos de la SW son minimalistas --como los denomina Eric Miller, líder de la Actividad de la Web Semántica del W3C-- y podrían resumirse en: a) Hacer sencillas las cosas sencillas y posibles las cosas complejas; b) No normalizar más de lo necesario y c) Crear sistemas cuya resultado sea mayor que la suma de las partes implicadas. Aunque la mayor utilidad cifrada en las tecnologías de la SW es mejorar o aumentar la eficacia en la recuperación de información en la Web, no es la única. La misma consistencia que implica esa "recuperación más inteligente", es decir el acceso a la información por conceptos y su representación en RDF, mejorará la accesibilidad de la información, al menos en dos sentidos: a través de la capacidad de los metadatos de anticiparse al contenido/condiciones de acceso de un objeto de información digital, y a través de la creación de perfil/es específico/s de aplicación de los metadatos a para incrementar cuestiones con implicaciones no sólo técnicas, sino también sociales como son la accesibilidad, la valoración de contenidos ([PICS](#)) o la creación de una Web de confianza (que retoma la idea de PGP, *Pretty Good Privacy*), cuestiones todas ellas, que descansan en tecnologías de la SW como RDF.

Por otra parte es importante destacar que, por diversos motivos, estamos en un buen momento para hablar de la Web Semántica y para profundizar y crear en su potencial. Algunas de las razones que justifican esta afirmación son, por ejemplo:

- a) El nivel de estabilidad y formalización de facto a través de las especificaciones de acceso público del W3C de los estándares vinculados al desarrollo de la SW. Desde el 10 de febrero de este año, todas las especificaciones del W3C vinculadas a la Web semántica se convirtieron en Recomendaciones ([seis relativas a RDF](#) y [seis especificaciones sobre OWL](#)), reafirmadas a principios de este mes, con la declaración de que tanto el [grupo de trabajo sobre RDF como el de OWL han completado con éxito sus productos finales](#).
- b) El nivel de difusión y desarrollo de conceptos, herramientas y tecnologías implicadas en la SW. Este nivel de difusión va desde el trabajo en listas especializadas y en el canal [#RDFIG](#), hasta las [actividades que ha desarrollado el W3C](#) en el contexto del proyecto SWAD-E o la proliferación de noticias en Weblogs que, de forma espontánea o formal constituyen un acervo de reflexiones sobre el tema, desde distintos puntos de vista.

En relación a la difusión de la SW debemos comentar que el mes de Junio de 2004 ha sido muy importante en este sentido en nuestro país. Por una parte se celebraron dos talleres sobre la Web Semántica en Madrid, el primero de ellos sobre "anotación de imágenes", albergado por el grupo de Ontologías de la UPM, los días 7 y 8 de junio (se puede [consultar la información sobre este taller](#), así como la discusiones que se mantuvieron a lo largo del mismo --[día 7](#) / [día 8](#)--); y el segundo, celebrado el 13 de junio por LaRed, y desarrollado completamente en español donde se planteó una [introducción a Web Semántica](#) y donde pudimos ver algunos de los desarrollos que se están llevando a cabo en este sentido en España y/o en español (también se puede acceder al fichero de logs con la discusión). Por otra parte, SEDIC, que ya ha demostrado su interés por estos temas, sobre todo a través del Grupo de Trabajo sobre Normalización para la Recuperación de Información en Internet ([NORMAWEB](#)), organizó una actividad de formación profesional titulada [Introducción a la Web Semántica para documentalistas](#), los días 24 y 25 de este mismo mes. En este curso, tratamos de poner de relieve sobre todo, la realidad de la Web Semántica en la actualidad. A modo de resumen podemos identificar cuatro de las aplicaciones actuales de la SW con mayor potencial de desarrollo en el trabajo documental:

- 1.- Descripción de recursos: Metadatos y codificación de esquemas de

metadatos en RDF. Esta podría ser identificada a priori como la principal función de los documentalistas, aplicando estándares como la [DCMI](#), codificando sus 15 elementos en RDF. La definición de ontologías y terminologías, o bien tesauros especializados para la organización y recuperación temática de la información, que se basan, bien directamente en RDF o a través de estándares específicos como Web Ontology Language ([OWL](#)) *Simple Knowledge Organization System* ([SKOS](#)) para la definición de tesauros.

- 2.- Descripción de imágenes. Esta es otra tarea importante desarrollada en los servicios de información. El ámbito de la descripción de imágenes compete a la descripción en sistemas de información, sistemas cuyo objeto de información digital por antonomasia es la representación fotográfica de un objeto dado (p. ej., información museística, información geoespacial, sistemas de fotografía digital de prensa, etc.) que utilizan el estándar para la descripción de contenidos, combinando otros espacios de nombre como FOAF, DC, etc. (algunos ejemplos son los desarrollados por [Libby Miller](#), [Morten Frederiksen](#), o el propio sistema de anotación de imágenes desarrollado por el W3C, [RDFPic](#)).
- 3.- La Agregación/sindicación de contenidos. Este es sin duda alguna y paradójicamente, ya que el estándar que utiliza ([RSS](#)) no ha sido desarrollado por el W3C, el ámbito más visible hoy en día de la aplicación de RDF a portales y bitácoras. Probablemente su visibilidad se deba a que, a lo largo de este año se han desarrollado múltiples herramientas de software para syndicar o agregar contenidos de noticias distribuidas (incluso la última versión del navegador Opera 7.5 incluye en su *suite* de programas una aplicación de este tipo que reciben genéricamente el nombre de *Newsfeed*) y desarrollar ficheros en RSS que describan las noticias de un determinado Website, se ha convertido en una nueva forma de visibilidad de la Web.

A estas aplicaciones básicas de la SW en Documentación tendríamos que añadir otras como la anotación de documentos con herramientas como [Annotea](#)/Amaya, [Annozilla](#), la gestión y descripción de marcadores y favoritos, donde algunos navegadores como Mozilla utilizan RDF para esta función, o las cada vez más de moda Redes sociales, tipo [eConozco](#), [Orkut](#), [Livejournal](#) que, a través de esquemas específicos como [FOAF](#) identifican y localizan las relaciones interpersonales.

Me gustaría terminar esta pequeña reflexión que estoy compartiendo con ustedes sobre la Web Semántica, citando, una vez más, a su principal mentor, Tim Berners-Lee, en su libro *Weaving the Web* (traducido al español por la editorial Siglo XXI en mayo de 2000 como *Tejiendo la red: el inventor del World Wide Web nos descubre su origen*): Si HTML y la Web hicieron que todos los documentos en línea parecieran un solo libro enorme, el RDF y los lenguajes de esquema o deducción harán que todos los datos del mundo parezcan una enorme base de datos. Esta es la segunda parte del sueño de Berners-Lee sobre la Web, donde no sólo colaboran las personas, sino que esa colaboración se extiende a los ordenadores, donde el ser humano, el documentalista, debe de aportar la inspiración y la intuición. Es incuestionable que la World Wide Web ha cambiado nuestras vidas y la sociedad moderna, que no en vano se denomina Sociedad de la Información. Ahora sólo nos resta esperar, calcular e incluso predecir, cómo va a cambiar la profesión del gestor de la información a raíz de la Web semántica, una Web "más bibliotecaria", donde las aportaciones de la experiencia acumulada por los profesionales de la información será un valor fundamental en la creación de sistemas robustos e interoperables basados en estas tecnologías.

CORESE: Conceptual Resource Search Engine	http://www.sop.inria.fr/acacia/soft/corese
DAML: DARPA Agent Markup Language	http://www.daml.org
DAML+OIL (Vid. DAML y OIL)	http://www.w3.org/TR/daml+oil-reference
DC: DCMi	http://www.dublincore.org
DC-Lib: Dublin Core Library Application Profile	http://dublincore.org/documents/library-application-profile
DCMES: Dublin Core Metadata Element Set (ISO 15836-2003)	http://www.niso.org/international/SC4/n515.pdf http://www.purl.org/metadata/dublin_core_elements
DCMI: Dublin Core Metadata Initiative	http://www.dublincore.org
DCMI-ES: Mirror en español de la DCMI	http://es.dublincore.org
FOAF: Friend Of A Friend	http://www.foaf-project.org http://xmlns.com/foaf/0.1
IMS: Global Learning Consortium	http://www.imspjproject.org
OAI: Open Archives Initiative	http://www.openarchives.org
OIL: Ontology Inference Layer	http://www.ontoknowledge.org/oil
OWL: Web Ontology Language	http://www.w3.org/TR/owl-features http://www.w3.org/2001/sw/WebOnt
PICs: Platform for Internet Content Selection	http://www.w3.org/PICS
RDF: Resource Description Framework	http://www.w3.org/RDF
RSS: RDF Site Summary	http://web.resource.org/rss/1.0
SKOS-Core: Simple Knowledge Organisation Systems	http://www.w3.org/2001/sw/Europe/reports/thes/1.0/guide
SOAP: Simple Object Access Protocol	http://www.w3.org/TR/SOAP
SVG: Scalable Vector Graphics	http://www.w3.org/Graphics/SVG
SW: Semantic Web Activity (W3C)	http://www.w3.org/2001/sw
SWAD-E: Semantic Web Advanced Development in Europe	http://www.w3.org/2001/sw/Europe
SWWS: Semantic Web Web Services	http://swws.semanticweb.org
UDDI: Universal Description, Discovery and Integration	http://www.uddi.org
W3C: World Wide Web Consortium	http://www.w3.org
WebOnt: Web Ontology Working Group (W3C)	http://www.w3.org/2001/sw/WebOnt
WS: Web Services Activity (W3C)	http://www.w3.org/2002/ws
WSDL: Web Services Description Language	http://www.w3.org/TR/wSDL
XML: eXtensible Markup Language	http://www.w3.org/XML
XTM: XML Topic Maps	http://www.topicmaps.org/xtm/1.0

> [Archivo de CLIP anteriores](#)