

Ponencia presentada en el 8° Congreso Nacional de Bibliotecología y Ciencias de la Información: "Bibliotecas y Educación en el Marco de la Globalización", organizado por la Asociación Colombiana de Bibliotecólogos y Documentalistas – ASCOLBI. Cartagena de Indias, Mayo 31 a Junio 2 de 2006.

## **RETOS PROFESIONALES EN LA NUEVA INTERNET: LA WEB SEMÁNTICA**

Wilmer Arturo Moyano Grimaldo  
Profesional en Ciencia de la Información – Bibliotecólogo  
[wilmer.moyano@rojaseberhard.com.co](mailto:wilmer.moyano@rojaseberhard.com.co)

Han pasado ya más de 15 años desde que la Internet se convirtió en la WWW, sin embargo los profesionales de la información han permanecido bastante ajenos a las bondades que esto conlleva. La presente ponencia desea exponer qué es la Web Semántica, indicando que el papel de los bibliotecólogos en la organización de la Web es un campo de acción que no es exclusivo de los profesionales en sistemas, y que por el contrario, es un lugar donde se pueden extraer importantes herramientas para la recuperación de información y para el apoyo a proyectos de carácter documental e investigativo.

### **Breve contexto histórico de la Internet y la Web**

Los orígenes de la Internet son recientes, con su creación se desarrollan muchas estructuras de organización del conocimiento que han permitido de cierta forma mejorar la calidad de vida de las personas, pero en otros casos a desmejorarla.

Si bien la Internet no surge como una idea de ser una biblioteca universal, y está aun muy lejos de serlo, ya es considerada por muchos como el lugar que a futuro proporcionará mas información para todos. Sin embargo, la idea de organizar el conocimiento en un todo no es algo reciente, existen contextos históricos que muestran que han existido varios intentos por conseguirlo, así como sistemas de clasificación del mismo, tal y como se puede ver en los trabajos de Aristóteles, Bacon, Melvil Dewey o Charles A. Cutter.

Uno de los principales intentos de organización del conocimiento lo desarrolló el señor Conrad Gesner, el cual es mejor conocido como el Padre de la Bibliografía. Entre el año 1545-1555 escribe una obra denominada "Bibliotheca Universalis" el cual era una obra en 4 volúmenes con el que pretendía agrupar toda la bibliografía existente en su tiempo. Aunque su obra ha sido olvidada por las nuevas generaciones, recientemente ha sido nuevamente retomada por el G7 para realizar algo similar con el apoyo de la Internet.

Los orígenes de la Internet son mas antiguos de lo que parece, sin embargo se toma su origen a partir del año 1969 cuando el proyecto ARPANET del sistema de Defensa de los Estados Unidos es creado, con el fin de tener un medio de comunicación efectivo para la guerra fría. Casi inmediatamente surge el primer proyecto de biblioteca digital, el llamado Proyecto Gutenberg, el cual aun se conserva vigente con una colección de mas de 18 mil títulos. Conforme el tiempo pasó surgieron muchas de las herramientas modernas que han permitido la evolución de la Internet a como es conocida hoy en día; en los años setenta del siglo XX surgen por ejemplo el correo electrónico y el Protocolo TCP/IP, en los años

ochenta surge la Internet como tal y se ven los primeros pasos de la World Wide Web, en los noventa se masifica y en la primera década del siglo XXI, surgen ideas para un mejor desempeño de la misma como lo son la Web Semántica. Es por todo esto que Tim O'reilly (2004) considera que ha habido tantos cambios en la Internet que estamos ya en la Web 2.0., donde ahora la red es mas abierta a los usuarios.

Muchas personas han permitido que la Internet como se conoce actualmente haya logrado surgir, pero vale la pena destacar principalmente a cuatro:

1. Vannevar Bush: quien en 1945, recién terminada la guerra, publica el artículo: "As we may think", donde expresó de manera casi profética su preocupación por la cantidad alarmante de información surgida hasta ese momento. Para organizar esto, propone la idea de un aparato para recuperar documentos a través de referencias cruzadas al que denominó "Memex"; la idea de este aparato era almacenar todo en microfilmes y que los criterios de búsqueda permitieran al lector recordar documentos con información similar, algo muy parecido a lo que años mas tardes haría la Internet con ayuda del hipertexto y los enlaces.

2. Douglas Engelbart: quien en la década de los sesenta del siglo XX propone la creación de un sistema de recuperación de información en ambiente de computadora al que denominó NLS (oNLine System). Al hacerlo de esta forma garantizaba el acceso a la información a muchos mas usuarios.

3. Ted Nelson: quien en 1965 durante la vigésima conferencia anual de la Association of Computer Machinery (ACM) propuso la creación de un modelo de interconexión de documentos electrónicos, al cual denominó "Hipertexto". Su creación se desarrolló a través de un proyecto propio llamado Xanadu, el cual aun existe. El hipertexto como tal es la base de lo que mas tarde se denominaría la WWW.

4. Tim Berners-Lee: en 1989 a partir de las teorías de todos los anteriores, crea la WWW, como un modelo hipertextual de recuperación de información en ambiente Web. A partir de este momento es que la Internet se convierte en lo que se conoce, y surgen nuevas herramientas para su evolución como el protocolo http o el lenguaje Html.

Cabe destacar que en todo este proceso de creación y organización de la Internet, ha sido muy poca la colaboración que ha habido por parte de profesionales en la organización de la información, y su tarea ha recaído casi exclusivamente en profesionales concedores de sistemas y tecnología.

### **De la Internet a la Web Semántica**

Hacia el año 1989, y basado en ideas ya preexistentes, Tim Berners-Lee, un ingeniero que trabajaba para la Organización Europea de Investigación Nuclear (CERN), tuvo la idea de diseñar un sistema basado en el hipertexto, con el fin de compartir información científica dentro de su organización de una manera mas eficiente. Con esta idea, su sistema estaba de alguna u otra forma emulando la concepción que tuvo Vannevar Bush en los 40s con su invento llamado "Memex". Es asi como Berners-Lee, junto a Robert Cailliau desarrollaron primer Browser de la historia, al cual denominaron World Wide Web.

El principal inconveniente con el que se encontraron fue que los formatos de los documentos no podían ser leídos por todos los computadores, navegadores y sistemas operativos del momento, de tal forma que diseñaron el lenguaje HTML, derivado del SGML y que es al que se le debe la forma en la cual los usuarios ven la Internet hoy en día y crearon el protocolo de red http. De esta forma nació la WWW como actualmente se conoce.

Para el año 1993, ya existían 50 servidores en el mundo que utilizaban la idea revolucionaria de Berners-Lee y en junio de ese año ya existían 130 sitios en Internet con esas características. El crecimiento fue tan sorprendente que para enero de 1997 la cifra de sitios en Internet alcanzaba unos 650 mil, lo que indica un crecimiento del 500 mil por ciento en 42 meses.<sup>1</sup>

Ante este crecimiento tan increíble, los problemas surgieron rápidamente, el lenguaje HTML inicial sobre el cual se escriben las páginas Web se convirtió en algo que cualquier persona podía manipular y acabó “degradándose” a un estándar sin normalización alguna, lo cual se manifestó en detalles como la guerra entre navegadores: Internet Explorer Vs. Netscape.

Por otra parte, ante la facilidad para crear páginas Web, miles de sitios vieron la luz con información de mala calidad y la Internet se convirtió en un monstruo incontrolable. Viendo esto, Berners-Lee, creó el consorcio W3C y dentro de sus objetivos, propuso empezar a corregir muchas de las falencias que tenía su WWW inicial. Por tal razón, es que se empieza a vislumbrar el concepto Web Semántica, como una alternativa de trabajo dentro de la Web actual, para lograr una mejor representación del conocimiento a través de esta.

Para Berners-Lee (2000), la Web Semántica se puede entender como

*“...una extensión de la Web ya existente, en donde la información se encuentra bien definida y entendible, mejorando la cooperación y mejorando la comunicación y cooperación entre el hombre y la computadora<sup>2</sup>...”*

Con esta idea se pretende construir una Web que sea mas acorde con las necesidades de los usuarios que la consultan. Para lograr todo esto, el creador de la WWW, considera que la Internet ideal debe tener dos alcances:

1. Si las máquinas fueron creadas por el hombre, deben estar a su servicio, por tal motivo debe ser muy rápida y de fácil acceso
2. Debe mejorar la interacción entre hombre-máquina, de tal manera que las computadoras puedan analizar mas fácilmente las necesidades de información del ser humano. Es aquí donde empieza a formarse el papel del profesional en información en la organización semántica de contenidos en la Web, y es donde se han perdido los límites entre los

---

<sup>1</sup> Op cit. Introducción a los metadatos. P. 11

<sup>2</sup> Cita textual: “The semantic web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation”

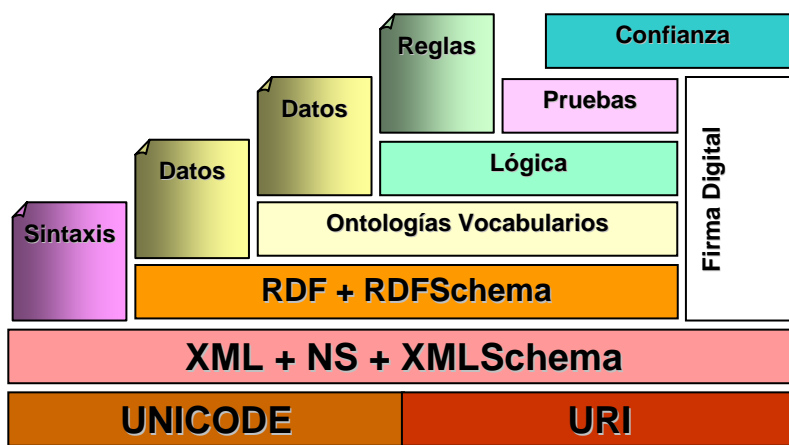
bibliotecólogos y las demás profesiones que de una u otra forma confluyen en a conformar la Sociedad del conocimiento, y la Web.

Sin embargo, es importante tener en cuenta que la Web Semántica parte de los postulados originales que tenía el creador de la WWW en el año 1989 por lo tanto, no es algo realmente nuevo aunque la metodología y las herramientas para su construcción si lo son.

### La semántica de la Web

Para entender hacia donde debe confluír la semántica en la Web y por qué es importante, se hace necesario entender a grosso modo la estructura de lo que sería la Web Semántica, de esta forma se podrá entender el rol del bibliotecólogo en la misma.

Si la Semántica tiene que ver con la significación de las palabras, en el caso de la Web debe estar dada principalmente por la creación de documentos digitales (objetos digitales) pero con significado, de tal forma que ya no se hable de información sino de conocimiento. Ante esto se hace evidente el papel que cumplirán los metadatos para conseguir este fin, pero no pueden funcionar por si solos, requieren de unas bases tecnológicas que se explican a continuación.



Como se aprecia en el grafico, la Web Semántica está conformada por varias capas, las superiores son las que dan el significado a los contenidos en tanto que las inferiores dan el soporte tecnológico y son la parte que el usuario final ve en la pantalla de su computador. El profesional en información está de una forma u otra vinculado a cada capa, pero su principal papel está en la organización de los contenidos dentro de todo el proceso.

En el caso del XML, éste da a los documentos la opción de un marcado semántico que permite recuperar la información contenida dentro de los mismos, ya que este lenguaje de marcado no muestra los atributos del documento sino las características del mismo, Ej. Titulo, párrafos, firmas, entre otras. Sin embargo para conseguir recuperar estos documentos, el lenguaje XML debe llevar asociados metadatos.

En vista que la recuperación por medio del XML y los metadatos no es suficientemente precisa, se crea una capa superior, la del RDF (Resource Description Framework) que es el marco donde se van a indicar los metadatos y la estructura semántica del recurso Web. Es

aquí donde inicia la Web Semántica, ya que es la primera capa con significado propio y que permite generar una primera relación entre el hombre y la máquina para que se entiendan. La idea del RDF es proporcionar un orden al XML de tal forma que la estructura del marcado sea coherente con lo que el ser humano realmente necesita. Esta capa es fundamental en la bibliotecología, porque también es de gran uso en el desarrollo de bibliotecas digitales a nivel mundial y es donde se ingresarán los metadatos, de acuerdo a su estructura.

Aunque el objetivo de esta ponencia no es el mostrar la estructura de la Web Semántica, ya que se encuentra mucha información en la Internet sobre este tema, sino mostrar las oportunidades que se presentan para los profesionales en información, es importante solamente indicar la estructura del mismo. De tal forma, el RDF tiene dos partes principales:

1. El modelo de datos: que es donde se ingresan los metadatos, para que adquieran la semántica necesaria, basado en una estructura de tripleta donde aparecen 3 componentes: recurso, propiedad y sentencia.
2. El esquema: que es la parte donde se definen los vocabularios y se define la estructura y las relaciones que llevará el RDF para cada caso particular.

Las capas superiores son las que le dan ya el aspecto que se quiere de Web Semántica, sin embargo son igualmente las capas donde menos han intervenido bibliotecólogos y en general profesionales en información. De estas capas, las más importantes son las ontologías; estas se pueden definir como una especificación explícita y formal de una conceptualización, tendiente a ser usada en un proceso basado en Inteligencia Artificial. Es por eso que el alcance máximo de la Web Semántica es el uso de agentes inteligentes para la recuperación de información basados en el desarrollo de ontologías. Esto puede parecer novedoso para muchos pero en realidad es un proceso que ya se ha desarrollado manualmente por bibliotecólogos en el pasado; por ejemplo el desarrollo de tesauros, o incluso la estructura de los sistemas de clasificación se comportan como ontologías muy primarias, sin embargo el proyecto de Berners-Lee está inundado de otros profesionales haciendo el trabajo de bibliotecólogos en el desarrollo de ontologías.

Por último toda la información de la WS, debe tener una validez que la dan aspectos como las pruebas del sistema y la firma digital, todo esto debe llevar a una Web que sea confiable para el usuario, que no genere dudas. Es por esto que a la capa más superior se le denomina la Web de la Confianza.

### **Opciones de trabajo desde la Ciencia de la Información**

La Web Semántica no puede entenderse como un cambio en el paradigma de la profesión, sino más bien como una oportunidad de demostrar la adaptabilidad y la vigencia de la misma. La esencia del quehacer de la profesión no cambia; la tecnología ha estado presente siempre en la evolución de la misma, por tal razón no es momento para asustarse por los sistemas y la Internet. Por otra parte hay que tener en cuenta que la tecnología solo es una herramienta que permite mejorar la calidad de vida del hombre no es un fin en sí misma.

De acuerdo a lo anterior, y si se siguen los alcances que propone Berners-Lee (2001) desde la Ciencia de la Información se debe tender a conseguir la colaboración hombre-máquina,

mediante el desarrollo de lenguajes documentales (Ontologías) que permitan llegar a esta meta y la unión de las teorías clásicas con las modernas.

Si se mira con detenimiento, existen unas labores que ya han tomado fuerza dentro de la Internet desde la profesión, las cuales pueden ser:

Desde la concepción de la Web

- Creación de servicios Web (portales temáticos)
- Estudios de Accesibilidad
- Conformación de objetos digitales
- Diseño de bibliotecas digitales
- Creación y gestión de contenidos en Bibliotecas digitales, museos digitales, archivos digitales

Desde los usuarios

- Capacitación de usuarios
- Alfabetización informacional
- Estudios de usabilidad
- Estudios de accesibilidad
- Servicios Web

Muchas de estas labores realmente no son diferentes de las tradicionales, solo cambia el ambiente en que se desarrollan. Sin embargo para conseguir la meta de la Web Semántica se requiere de una labor que es mucho mas explicita y es quizás la mas abandonada: la organización y recuperación del conocimiento.

Dentro de la organización de la Web, a partir de las ontologías se pueden tomar dos caminos:

1. Creación de nuevas ontologías partiendo de cero
2. Creación de nuevas ontologías basadas en lo que ya existe (mapeado).

La segunda opción es la mas viable y es donde ya profesionales en información han empezado a trabajar. Se pueden encontrar ejemplos como la creación de ontologías basadas en los sistemas de clasificación tradicionales, Sistema de Clasificación Decimal Dewey, Clasificación Decimal Universal, Clasificación Colonada o Clasificación de la Biblioteca del Congreso de los Estados Unidos. Un ejemplo de esto se puede ver en el trabajo que Saded y Chaudhry (2002) realizaron, al crear una taxonomía en el área de ciencia de los computadores mediante el mapeo del tesoro Web de la IEEE con los índices del SCDD.

Existen además otras opciones de combinación de la profesión con las tecnologías de la Web Semántica, Stefan Gradmann (2005) propone por ejemplo la unión del RDF con los Requisitos Funcionales de los Registros Bibliográficos (FRBR) con el fin de conseguir que los catálogos de las bibliotecas sean mas visibles en la Web. Por su parte, Wallis (2004) indica la conformación de una ontología basada en la estructura Entidad-relación del FRBR y el uso de RDF y los lenguajes para ontologías DAML+OIL para dar mas semántica al catálogo de su biblioteca. Un ejemplo práctico de esta unión es la que se puede ver en el software bibliográfico Virtua de la empresa VTLIS.

Finalmente es importante recalcar que las tecnologías de la Web Semántica se pueden usar para la implementación de Bibliotecas Digitales en todo el mundo, la Networked Digital Library of Theses and Dissertations usa por ejemplo el RDF para la implantación de su núcleo de metadatos; esto debido a que las bibliotecas digitales al encontrarse en un ambiente Web también contribuyen y se benefician de las estructuras de la Web Semántica.

### **Conclusiones**

A manera de conclusiones se puede afirmar que:

- La Web Semántica no es una nueva Web es la potenciación de la ya existente hacia una forma mas semántica de la misma.
- La posibilidad de llegar a esta meta es aún remota, porque hace falta mas trabajo interdisciplinario. Se requiere igual de mucha mas investigación documental desde las universidades, especialmente de las que tienen carreras en Ciencia de la Información.
- Es un proyecto que implica apoyo de los gobiernos, porque estos son los que garantizarán las políticas nacionales e internacionales de acceso a la Internet.
- Es un campo de mucho trabajo e investigación por parte de bibliotecarios y documentalistas, pero si no lo hacemos otros lo harán. Ej. Grupo de investigación de la Universidad del Valle
- No es un cambio en la forma de ejercer la profesión, es una adaptación de lo que se ha venido haciendo desde siempre. (Almacenamiento, administración, organización y recuperación de información y conocimiento).

## Referencias bibliográficas

BERNERS-LEE, Tim. Weaving the Web: the original design of the World Wide Web by it's inventor. New York: Harper Collins, 2001 246 p.

BERNERS-LEE, Tim; Hendler, James y LASSILA, Ora. The semantic Web: a new form of web content that is meaningful to computers will unleash a revolution of new possibilities. En: Scientific American, No. 501, mayo 2001 [En línea] <http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21> [Consultado en febrero 01 de 2004]

CARDOZO F. Lesly e YZQUIERDO H. Jesús R. Diseño conceptual de un portal Web Semántico para el área de la salud en Venezuela. [Trabajo de grado de Licenciatura en Bibliotecología]. Caracas, Universidad Central de Venezuela. Escuela de Bibliotecología y Archivología, 2005. 145 p.

CARTER, Jason. Dewey's Ontology: the underlying order of the Dewey Decimal Classification System [en línea] <http://www.pages.drexel.edu/~jmc63/FinalPaper.htm> [Consultado en mayo 12 de 2006]

CHOWDHURY, Suddata y CHOWDHURY G.G.. Using DDC to create a visual knowledge map as an aid to online information retrieval. [en línea] En: ISKO8: Knowledge organization and the Global Information Society. ISKO. 2004 [http://www.cis.strath.ac.uk/research/publications/papers/strath\\_cis\\_publication\\_333.pdf](http://www.cis.strath.ac.uk/research/publications/papers/strath_cis_publication_333.pdf) [Consultado en mayo 15 de 2006]

GRADMANN, Stefan. rdfs:frbr-. Towards an Implementation Model. for Library Catalogs. Using Semantic Web Technology. En: Cataloging & Classification Quarterly. Vol. 39, No. 3/4, 2005. P. 65-75.

HAVU, Sirkka. Bibliotheca universales: how to organize chaos?. En: Satellite meeting to the 71st world library and information congress (71 : 2005 : Järvenpää, Finland) [en línea] <http://www.kaapeli.fi/~fla/frbr05/GesnerByHavu2005.pdf> [Consultado en octubre 29 de 2005].

MENDEZ, Eva. La Web Semántica: una web "más bibliotecaria". [en línea]. En: Boletín Clip, No. 41, 2004. [http://www.sedic.es/p\\_boletinclip41\\_confirma.htm](http://www.sedic.es/p_boletinclip41_confirma.htm) [Consultado en mayo 10 de 2006]

O'REILLY, Tim. What is Web 2.0: design patterns and business models for the next generation of software. [en línea] En: O'reilly. 2005. <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html> [Consultado en mayo 01 de 2006]

SAEED, Hamid; CHAUDHRY, A.S. Using Dewey Decimal Classification Scheme (DDC) for building taxonomies for knowledge organisation. En: Journal of Documentation, Vol. 58., No. 5. 2002. p. 575-583

VIZINE-GOETZ, Diane. OCLC investigates using classification tools to organize Internet data [en línea] OCLC investigates using classification tools to organize Internet data <http://digitalarchive.oclc.org/da/ViewObject.jsp?fileid=0000003965:000000105474&reqid=4496#investigates> [Consultado en mayo 01 de 2006]

VIZINE-GOETZ, Diane. Using library classification schemes for internet resources [En línea]. [Dublin, OH]: OCLC, Internet Cataloging Project Colloquium, 1999. <http://staff.oclc.org/~vizine/InterCat/vizine-goetz.htm> [Consultado en Abril 10 de 2006]

WALLIS, Jillian C. An argument for a Semantic Web based FRBR union catalogue [en línea]. [Los Angeles]: UCLA, 2004 <http://www.moebiusstrip.org/277/final.html> [Consultado en mayo 10 de 2006]