

# La Web Semántica<sup>1</sup>

**Por: Lluís Codina y Cristòfol Rovira**

Universidad Pompeu Fabra

Departamento de Periodismo y de Comunicación Audiovisual

Instituto Universitario de Lingüística Aplicada

[www.semanticweb.net](http://www.semanticweb.net)

[www.hipertext.net](http://www.hipertext.net)

2006

---

## Resumen

La Web semántica es un proyecto a corto, medio y largo plazo del organismo de regulación más importante del mundo en relación a Internet: el World Wide Web Consortium (W3C a partir de ahora). El proyecto de la Web semántica incluye transformaciones que ya están afectando a los ámbitos de la creación, edición y publicación de páginas y sitios Web<sup>2</sup> y que seguirán teniendo una importancia creciente en el futuro.

Este capítulo expondrá los conceptos e ideas más importantes relacionadas con la Web semántica, siempre con el telón de fondo de los intereses de la Biblioteconomía-Documentación.

## Sumario

<b>1 INTRODUCCIÓN.....</b>	<b>2</b>
<b>2 LA WEB SINTÁCTICA.....</b>	<b>5</b>
<b>3 BÚSQUEDA BASADA EN CADENAS DE CARACTERES.....</b>	<b>5</b>
<b>4 LA WEB SEMÁNTICA.....</b>	<b>7</b>
<b>5 INFRAESTRUCTURA DE LA WEB SEMÁNTICA.....</b>	<b>9</b>
5.1 XML.....	11
5.2 METADATOS Y RDF .....	14
5.3 ONTOLOGÍAS.....	21
5.4 OWL.....	23
<b>6 PROBLEMAS EN EL PARAÍSO.....</b>	<b>24</b>
<b>7 MIGRACIÓN: UN DISEÑO EXPERIMENTAL.....</b>	<b>24</b>
7.1 ANÁLISIS.....	25
<b>8 CONCLUSIONES.....</b>	<b>28</b>
<b>9 BIBLIOGRAFÍA.....</b>	<b>30</b>
<b>10 ANEXO: GRUPOS DE SEDES WEB ANALIZADOS.....</b>	<b>32</b>

---

<sup>1</sup> Capítulo del libro *Tendencias en documentación digital* (Trea, 2006). Forma recomendada de citación: <Lluís Codina, Cristòfol Rovira. "La Web semántica". En: Jesús Tramullas (coord.). *Tendencias en documentación Digital*. Guijón: Trea, 2006, p. 9-54>

<sup>2</sup> Escribiremos Web con mayúscula cuando nos refiramos al conjunto de la Web (es decir a la World Wide Web); mientras que escribiremos web con minúscula para referirnos a un solo sitio o una sola página web.

# 1 Introducción

El W3C ([www.w3.org](http://www.w3.org)) es el organismo que regula aspectos esenciales de la Web tales como el lenguaje (X)HTML con el cual se crean las páginas y los sitios web. Puede decirse que es, con mucha diferencia, el organismo de normalización más importante de Internet, siendo su director el propio fundador de la Web, Tim Berners-Lee, por lo que sus recomendaciones, que tienen carácter normalizador, poseen un gran prestigio y una enorme influencia. La Web semántica es el proyecto del W3C para transformar la Web en la Web de las próximas décadas. Ante todo, veamos la definición oficial de la Web semántica según el W3C:

La Web semántica proporciona un marco común que permite que los datos sean compartidos y reutilizados a través de aplicaciones, empresas y fronteras comunitarias. Es un esfuerzo colaborativo liderado por el W3C con la participación de un gran número de investigadores y socios industriales. Está basado en *Resource Description Framework* (RDF) e integra una variedad de aplicaciones utilizando XML para la sintaxis y URI para las denominaciones ([www.w3.org/2001/sw/](http://www.w3.org/2001/sw/))

Dos breves apuntes sobre la definición anterior. En primer lugar, parece un tanto críptica, tal como acostumbra a ser, de hecho, las definiciones del W3C. Lo segundo que corresponde señalar es que la Web semántica no (aún) es una realidad. De acuerdo con las estimaciones del W3C, el despliegue total de la Web semántica puede prolongarse más allá del año 2010.

Sin embargo, la Web semántica ya está entre nosotros de diversas formas. En primer lugar, bajo la forma de una auténtica idea-fuerza, en el sentido de que es una idea que ya ha sido capaz de movilizar energías (e ilusiones) y que, sin duda no dejará de arrojar resultados positivos durante los próximos años. En segundo lugar, aportando nuevos estándares que ya son de uso habitual (como el lenguaje XML) e influenciando en el desarrollo de la nueva generación de navegadores y editores de páginas web.

En todo caso, volviendo a su definición, en el proyecto de la Web semántica conviven dos grandes visiones o dos grandes ideas-fuerza cuya confluencia a veces dificulta su interpretación. Por este motivo, nosotros proponemos dos definiciones separadas (que se pueden complementar) de la Web semántica:

*Definición 1. La visión de la Inteligencia Artificial:* La Web semántica es un conjunto de iniciativas destinadas a promover una futura Web cuyas páginas estén organizadas, estructuradas y codificadas de tal manera que los ordenadores sean capaces de efectuar inferencias y razonar a partir de sus contenidos.

*Definición 2. La visión del procesamiento robusto:* La Web semántica es un conjunto de iniciativas destinadas a convertir la World Wide Web en una gran base de datos capaz de soportar un procesamiento sistemático y consistente de la información.

Lo que intenta poner en evidencia la primera definición es la visión o la idea-fuerza presente en el proyecto de la Web semántica que proviene de la Inteligencia Artificial (IA a partir de ahora). Es útil recordar que, históricamente, en el campo de la IA se han manejado dos hipótesis: las denominadas hipótesis fuerte y débil. La hipótesis débil sostiene que es posible conseguir ordenadores con inteligencia *simulada* y con diversos grados de éxito dependiendo del contexto. La hipótesis fuerte afirma que los ordenadores pueden alcanzar inteligencia *real* e *indiferenciable* de la humana (Penrose, 1991; Copeland, 1996).

Es evidente que los ordenadores actuales no son capaces de razonar ni de realizar inferencias en un modo similar al de los seres humanos, y tras varias décadas de investigación en IA, ni tan solo hay atisbos sobre qué clase de cambio de paradigma en la computación podría conducir en el futuro, aunque solo fuera hipotéticamente, a dotar de inteligencia real a las máquinas. Por tanto, debemos dejar claro que la clase de “razonamientos” que puede esperarse que sean capaces de realizar los ordenadores en el futuro sería, en el mejor de los casos, una *simulación* de razonamiento como la que postula la versión de la hipótesis débil de la IA.

Veamos ahora la definición 2 vinculada a la visión del procesamiento robusto. Lo que separa a un conjunto de documentos con información no estructurada, y por tanto difícil de procesar y de explotar su contenido respecto de un conjunto de registros de una base de datos es la suma de tratamiento sistemático + metadatos propia de estos últimos (y ausente en los primeros).

Recordemos que la creación de una típica base de datos documental consiste en definir un grupo de campos, lo que equivaldría en nuestro caso a definir un conjunto de etiquetas como <autor>, <título>, etc., para marcar sistemáticamente en cada documento de la base de datos la información que en el documento original aparece sin ninguna identificación explícita. El segundo paso consistirá en vincular cada documento con metadatos mediante etiquetas del estilo <clasificación>, <tipo de documento>, <descriptores>, <fecha de creación>, etc. (Abadal, Codina, 2005).

Una vez tenemos lo anterior, hemos pasado de información desestructurada a información sistematizada en la que cada línea de texto, cada párrafo o cada grupo de párrafos forma parte de un campo y está vinculado a un conjunto de metadatos. A partir de aquí será sencillo conseguir que la base de datos simule una cierta inteligencia de la que carecen en estos momentos los motores de búsqueda, ya que será capaz de responder a preguntas que actualmente no puede responder un motor de búsqueda. Por ejemplo, en la actualidad no existe forma de pedir a un motor de búsqueda que busque documentos donde la palabra Eco se refiera al nombre de un autor y no a un fenómeno acústico. En cambio, en una base de datos documental es una operación tan trivial que nos pasa absolutamente desapercibida.

Es a esta clase de procesamiento sistemático (predecible) y consistente a la que nos queremos referir con la expresión de procesamiento robusto.

Ahora bien, dada esta dicotomía, ¿hay algún elemento que nos permita unificar o al menos articular las dos visiones? La respuesta, al menos en nuestra opinión es que sí. Si observamos los elementos de infraestructura en los que confía la visión de la IA, vemos que son en parte los mismos que se requieren para crear una base de datos, es decir los mismos de la visión del procesamiento robusto.

En primer lugar, la visión de la AI requiere páginas codificadas de forma consistente, es decir, sin ambigüedad ni contradicciones; pero esto es exactamente lo que proporciona la estructurada basada en campos propia de una base de datos. En segundo lugar, la AI requiere una capa de metadatos que contenga declaraciones sobre las propiedades de los sitios web. Sucede que la asociación sistemática de metadatos a cada documento es lo que corresponde a la práctica de la indización, catalogación, categorización, etc., tan característica de las bases de datos en general, pero muy en particular, de las bases de datos documentales.

Lo que separa a ambas visiones es lo siguiente: la primera idea-fuerza es claramente visionaria, para bien y para mal, al confiar en obtener como resultado ordenadores capaces de razonar. Para bien porque sin duda a veces se requieren ideas visionarias para abrir nuevos caminos o para salir de una situación estancada. Para mal, porque a veces las ideas visionarias, al ignorar los hechos más elementales malgastan grandes esfuerzos.

La segunda visión, la del procesamiento robusto, está mucho más pegada al terreno. Es solvente, porque se basa en elementos bien probados en el procesamiento de la información, y esa es su gran virtud. Su problema es que carece de la capacidad de fascinación de la primera. Es posible que, si el proyecto de la Web semántica se hubiera limitado a esta segunda visión, nunca hubiera trascendido de las páginas de las revistas especializadas.

El proyecto de la Web semántica se enfrenta a retos cualquiera que sea la visión adoptada: nunca se había intentado aplicar la IA a un entorno abierto y descentralizada como es la Web. Tradicionalmente, la IA se había aplicado a dominios del conocimiento y conjuntos de datos bien diferenciados. El modelo clásico son los sistemas expertos, que siempre se limitan a un dominio y a un conjunto de datos restringido. Un ejemplo, es Dendral, un sistema experto para el análisis químico, o Mycin, un sistema experto que ayuda a diagnosticar enfermedades infecciosas de la sangre.

Por otro lado, la visión del procesamiento robusto también se enfrenta a retos. Las bases de datos funcionan bien porque, al igual que los sistemas expertos (aunque en un sentido distinto) se limitan a una colección de documentos bien delimitada, aunque sea enorme (pensemos por ejemplo en los millones de registros de Medline o Eric en los cientos de millones de documentos en texto completo de Lexis-Nexis). Es cierto que la Web semántica sería equiparable a una base de datos distribuida como las que ya existen actualmente. El problema es que no existen precedentes, ni mucho menos, de una base de datos distribuida con las dimensiones de la Web, y aún menos una base de datos distribuida que no cuenta con ninguna clase de mecanismo o procedimiento de coordinación entre los componentes de esa base.

En todo caso, para entender mejor lo que significa la este proyecto sin duda es útil considerar cómo es la Web actual, a la cual podemos denominar, por oposición a la Web semántica, la Web sintáctica.

## 2 La Web sintáctica

Las páginas HTML actuales disponen de etiquetas tales como h1, h2, etc., para marcar la importancia relativa cada sección de la página: en concreto, la etiqueta h1 está destinada a marcar el título principal de la página, mientras que h2, h3, etc., representan a su vez los títulos de las secciones de segundo, de tercer nivel, etc. Otro ejemplo, son las etiquetas para otorgar énfasis al texto, como cite para señalar la transcripción literal de un texto.

HTML, por tanto, aporta algunas etiquetas con valor estructural o funcional, mientras que otras etiquetas, como <b>, <i>, etc., sirven en cambio únicamente para señalar elementos gráficos; en concreto, indican al navegador que el texto que aparece entre los elementos <b> y </b> deben ser mostrados en negrita, mientras que el texto que aparece entre <i> e </i> debe ser mostrado en cursiva.

El problema con esta codificación es doble: no solamente carece de cualquier interpretación semántica, sino que, además, sus etiquetas son susceptibles de uso inadecuado: algunas páginas web contienen los elementos h1, h2, etc., intercalados de forma contraria al nivel estructural que representan, por ejemplo, puede aparecer un elemento h1 después de un elemento h2, para conseguir el efecto tipográfico asociado por el navegador con la etiqueta (negrita y un cuerpo más grande). También puede suceder exactamente lo contrario, es decir, que el título principal y los títulos de las secciones carezcan de la etiqueta correspondiente y, en su lugar, el autor de la página haya intentado marcar su importancia mediante atributos de formato (como negrita o cursiva y distintos cuerpo de letra) en lugar de estructurales (como h1, h2, etc.).

El resultado es una Web donde la codificación de las páginas, además de poseer un nulo valor semántico (una de las pocas excepciones es la etiqueta <title> de la sección de cabecera de la página), se puede utilizar de forma contraria a su función, por lo que ni tan solo son fiables los indicios que podrían proporcionar las etiquetas funcionales. En tales condiciones, la interpretación semántica certera de la página es imposible por parte de analizadores automáticos. Como resultado, las páginas web tienen semántica únicamente para los seres humanos.

En efecto, tal como se codifican las páginas web actuales, principalmente mediante el lenguaje HTML, tienen muy poco sentido para las máquinas. En efecto, si vemos el código fuente de una página web actual, encontramos, por ejemplo, un trozo de código como el siguiente:

```
...  
<b><i>Superar la brecha digital</i></b>  
...
```

cuando el ordenador lo interprete, a través del programa navegador, aparecerá como un texto en negrita y cursiva, como éste:

```
...  
Superar la brecha digital  
...
```

## 3 Búsqueda basada en cadenas de caracteres

Con lo anterior se acaba una buena parte de lo que es capaz de hacer un ordenador con las páginas HTML. Pero, como saben bien informáticos y documentalistas, otra cosa que pueden hacer los ordenadores es construir índices con las palabras que aparecen en las páginas web. Después cuando alguien envía una pregunta a un motor de búsqueda, lo que hace este último es

comparar las palabras de la pregunta con las palabras de su índice. Por ejemplo, supongamos que el responsable de un programa de gobierno sobre el problema de la brecha digital decide indagar en Internet para ver si encuentra estudios o informes sobre la brecha digital.

Supongamos que accede a Google y entra la siguiente pregunta: "brecha digital". Lo que hará Google es comparar las palabras de su pregunta, con las palabras de su índice. Si encuentra un documento que tenga la "brecha digital", lo devolverá como respuesta. Esto es casi todo lo que pueden hacer los ordenadores que tenga que ver con procesamiento de información en páginas web.

Con estas limitaciones, aunque la búsqueda en Internet está repleta de satisfacciones (no es difícil encontrar cosas valiosas en la Web con ayuda de los motores de búsqueda), también provoca muchas frustraciones. Si alguien busca por "caballos", no encontrará nada que trate sobre "yeguas" o sobre "potros". Si alguien busca sobre cómo evitar la guerra, no encontrará un documento sobre cómo conseguir la paz, por la simple razón de que las cadenas de caracteres no coinciden. En una búsqueda basada en la palabra clave "depresión" el ordenador no tiene forma de saber si buscamos documentos sobre geografía, sobre psicología o sobre el clima.

Quizás lo peor, con todo es la imposibilidad de precisar géneros documentales (artículos de revista vs. entradas de diccionario, por ejemplo) o puntos de vista en una búsqueda. Por ejemplo, una búsqueda sobre "Pentium 4" arrojará una buena cantidad de páginas de comercio electrónico. Si eso es lo que quiere el usuario, es decir, encontrar tiendas de informática en Internet, no hay ningún problema. Pero, si lo que quiere es encontrar documentación técnica sobre microprocesadores de la marca Pentium 4 así como análisis y comparaciones con otras marcas, por ejemplo, con AMD, probablemente deba dar la búsqueda por imposible dada la supremacía de las páginas web dedicadas al comercio electrónico en las listas de resultados de los motores de búsqueda.

Hay otras muchas cosas que no podemos expresar. Antes de pasar a la siguiente, que merece una cierta atención, tampoco es posible precisar lo que podemos llamar la granularidad de la respuesta. Si buscamos información sobre un país, por ejemplo, Botswana, no es posible indicar si queremos una síntesis breve de sus condiciones de vida sociales y de su forma de gobierno o un estudio detallado de economía con cientos de estadísticas sobre todas las dimensiones económicas y políticas de Botswana (a través de las cuales, sin duda, si pudiéramos dedicar el número de horas suficientes tal vez podamos deducir lo anterior).

Tampoco resulta posible expresar puntos de vista. Por ejemplo, supongamos la necesidad de información que denominaremos *N1*, y que consistente en que necesitamos encontrar información sobre aplicaciones de la Web semántica a la documentación. Hasta ahora, lo único que puedo hacer para expresar una necesidad de información como *N1* es construir una ecuación booleana, que denominaremos *E1*, como esta: [Web semántica AND documentación]. El problema es que la ecuación *E1* serviría a la vez para todas estas necesidades de información distintas, y de las cuales posiblemente solamente una de ellas es la requerida por el usuario

- i) *N1: Web semántica aplicada a la documentación.* Por ejemplo, RDF para representar lenguajes documentales
- ii) *N2: Documentación "de" la Web semántica.* Por ejemplo, el tutorial de un programa para editar archivos en formato XML
- iii) *N3: Documentación "sobre" la Web semántica.* Por ejemplo, la página oficial del World Wide Web Consortium sobre la Web semántica
- iv) *N4: Documentación aplicada a la Web semántica.* Por ejemplo, bibliotecas, archivos o centros de documentación dedicados a la Web semántica

Si tomamos como ejemplo a los lectores interesados en este capítulo, resulta que solamente los documentos que responden a la necesidad de información *N1* resultarán de utilidad. Un motor de búsqueda como Google, no obstante, devolverá como respuesta de forma indiscriminada documentos correspondientes a las necesidades de información *N1*, *N2*, *N3* y *N4* por la simple razón de que no existe ningún procedimiento para especificar los objetivos de la búsqueda, ni el género o la clase de documentos solicitados ni tampoco el punto de vista concreto que necesita el usuario.

## 4 La Web semántica

La Web semántica puede ser la respuesta a los problemas anteriores, aunque el proyecto intenta ir mucho más allá. Comencemos por señalar que, en la Web semántica, en lugar búsquedas por comparación de cadenas de caracteres, se espera que los sistemas de información sean capaces de buscar por conceptos.

Si buscamos por caracteres, las palabras de la pregunta y las palabras del documento (o del índice de documentos) deben coincidir letra a letra. En cambio, si buscamos por conceptos, lo de menos es la palabra. Lo importante es el concepto. Esto suena a inteligencia artificial. Por tanto, aunque existe una cierta resistencia a llamarlo así, con la Web semántica se está buscando el mismo objetivo que la IA, a saber, que los ordenadores entiendan que un documento sobre "equinos" puede ser muy relevante para una necesidad de información sobre "caballos", y que conceptualmente las preguntas "¿es posible parar la guerra?" y "¿es posible alcanzar la paz?" son en realidad la misma pregunta.

Por tanto, entre los objetivos de la Web semántica se encuentra la posibilidad de que sea posible sostener una interacción entre un usuario y un agente de software mediante el cual el primero pueda ir expresando y perfilando sin ambigüedad puntos como los siguientes: objetivos de la búsqueda, géneros documentales pertinentes, punto de vista, granularidad esperada en la respuesta, etc. A partir de aquí, se espera que el agente de software sea capaz de elaborar una estrategia de búsqueda según su propia iniciativa (la del agente de software) que involucre el uso de lenguajes documentales, metadatos y ontologías para responder con eficacia y rapidez al usuario.

Se espera igualmente que los ordenadores puedan desarrollar tareas de gestión que requieran interpretar información y tomar decisiones adaptándolas al contexto. El mejor ejemplo de este tipo de tareas lo proporcionó el propio Tim Berners-Lee junto con dos de sus colaboradores en el año 2001 (Berners-Lee, Hendler y Lassila, 2001: 39). En este artículo, los autores mencionados explicaban el caso de unos usuarios del futuro (para cuando se escribió ese artículo se estaba pensando en unos 5 años vista, por la tanto ya debería ser posible) podrían encargar a un agente de software el establecimiento de citas para sesiones de fisioterapia. Vale la pena reproducir lo que señalaba el citado artículo:

Lucía, desde la consulta del médico, dio instrucciones a su *agente* de la Red<sup>3</sup> semántica mediante su navegador portátil. Al cabo de unos instantes, el agente había obtenido del agente del médico la información necesaria sobre el **tratamiento prescrito**, había consultado varias listas de **profesionales**, y verificado cuáles pertenecían al **seguro médico en un radio de 30 kilómetros del domicilio** de ésta, y habían recibido **valoración de excelente o muy bueno** por servicios de evaluación de calidad dignos de confianza. El agente empezó entonces a buscar concordancias entre las **horas de cita previa** de que disponían estos profesionales (proporcionados por los agentes de cada uno de ellos desde

---

<sup>3</sup> Seguimos la traducción aparecida en castellano en el número de *Investigación y Ciencia* de Junio de 2001. Red en este caso es un sinónimo de World Wide Web (en general, Red en mayúscula se refiere a la Web)

sus sitios en la Red) y los escasos huecos con que contaban los atareados Germán y Lucía. (Berners-Lee, Hendler y Lassila, 2001: 39).

Nosotros hemos marcado en cursiva la palabra agente de la primer línea. El término agente en todo el párrafo citado (no solamente en la primera línea) se refiere a un programa informático con autonomía para realizar acciones e incluso adoptar decisiones (aunque sean luego revisadas por el usuario) inteligentes.

En cambio, los términos en negrita están así mismo en el original. Respecto a ellos, señalan los autores: “Las palabras clave resaltadas denotan términos cuya semántica [...] le fue definida al agente a través de la Red semántica”. Es decir, de acuerdo con los autores, términos como “tratamiento prescrito”, “profesionales”, “domicilio”, etc., serán entendidos, es decir, tendrán valor semántico, para el agente de software.

Presentada en estos términos, no cabe más remedio que concluir que la Web semántica incluye, como ya hemos destacada antes un objetivo al que la informática ha denominado hasta ahora Inteligencia Artificial. A continuación, siempre en el mismo artículo, puede leerse lo siguiente:

La Red semántica aportará estructura al contenido significativo de las páginas de la Red; creará un ambiente en el que los agentes informáticos, que irán visitando una página tras otra, realizarán con presteza tareas complejas por encargo de los usuarios del sistema. Llegado a la página de la clínica, tal agente no sólo *sabr*á que la página contiene términos clave como “tratamiento, medicina, fisioterapia”, sino que el doctor Martín **consulta** en esta **clínica** los **lunes, miércoles y viernes**, que el guión acepta **intervalos de fechas** en el formato **dd-mm-aa** y responde con **horas de cita**. Podrá “enterarse” de todo ello sin necesidad de poseer una inteligencia artificial”. El contenido semántico quedó codificado en la página de la clínica cuando sus administradores le dieron forma [...]. (Berners-Lee, Hendler y Lassila, 2001: 39).

En el párrafo precedente vemos que, después de enumerar tareas propias de la Inteligencia Artificial, los autores indican claramente contradictoria: “todo ello sin necesidad de poseer una inteligencia artificial”. Esta declaración solamente se explica por el deseo expreso del W3C de distanciarse del aura de fracaso de una parte de la Inteligencia Artificial que, entre los años 50 y 80 proclamó día sí y día también que las máquinas inteligentes estaban a la vuelta de la esquina sin que tal cosa haya sido cumplida ni siquiera 50 años después de las primeras promesas.

A pesar de todo, un sencillo test con Google nos indica que, en febrero del 2006, una pregunta sobre “Inteligencia Artificial y Web semántica” arrojó más de medio millón de resultados que relacionan ambas cosas (recordemos que, para Google, la simple yuxtaposición corresponde a un AND booleano).



Figura 1: Una consulta en Google sobre Inteligencia Artificial y Web semántica

## 5 Infraestructura de la Web semántica

Los medios con los cuales se persiguen los objetivos de la Web semántica que hemos presentado anteriormente son, a grandes rasgos, los siguientes: en primer lugar, mediante una codificación de páginas en la cual las etiquetas tengan, precisamente, carga semántica. Este apartado corresponde al estándar denominado XML (*eXtensible Markup Language*).

En segundo lugar, aportando descripciones (metadatos) de las páginas y sitios web con un formato que sea compatible con la estructura general de la Web y con diversas categorías de páginas e interoperable entre distintos sistemas informáticos. De este se ocupa la norma RDF (*Resource Description Language*).

En tercer lugar, mediante un sistema de ontologías que permitan especificar conceptos de los diversos dominios del conocimiento mediante el uso de un lenguaje fuertemente basado en lógica simbólica y susceptible, por tanto, de ser eventualmente interpretado por un ordenador. De este aspecto se ocupa el denominado *OWL Web Ontology Language* (OWL).

No obstante, el proyecto de la Web semántica está formado por una auténtica sopa de letras, dada la diversidad de normas, protocolos, lenguajes y especificaciones involucradas. De hecho, existe un famoso diagrama debido a Tim Berners-Lee, de gran capacidad expresiva, que pretende abarcar la totalidad del proyecto mediante una metáfora de capas que comprende 7 niveles y que reproducimos a continuación:

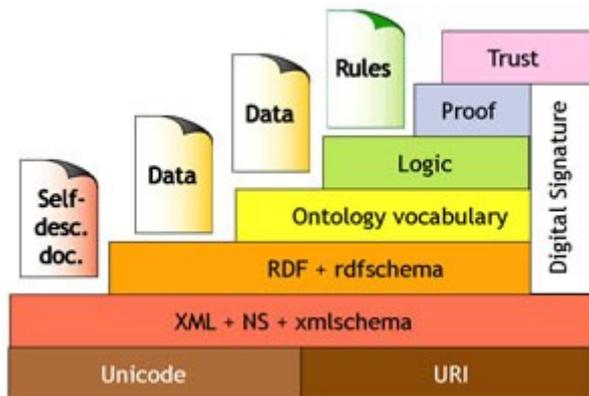


Figura 2: La Web semántica vista como un sistema de capas (fuente: Tim Berners-Lee. <http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide11-0.html>)

En la tabla siguiente presentamos más detalle el significado de las capas del gráfico de la Figura 2.

**Tabla 2: Los 7 niveles de la Web Semántica**

<b>1</b> <i>Unicode + URI</i>	<i>Unicode</i> es un sistema internacional estándar que proporciona un número único para cada carácter, sin importar la plataforma ni el programa. Esto permite representar caracteres de cualquier idioma con una codificación unificada. <i>Uniform Resource Identifier</i> (URI) es un sistema de direccionamiento e identificación de recursos. El sistema que usamos actualmente para acceder a los recursos de la Web (URL) es una parte de URI.
<b>2</b> <i>XML+NS+XML SCHEMA</i>	<i>eXtended Markup Language</i> (XML) es un sistema que permite definir lenguajes de marcas para usos específicos. <i>Name Spaces</i> (NS) permite combinar diversos lenguajes de marcado creados con XML en un mismo documento. <i>XML Schema</i> sirve para definir tipos de documentos complejos en los que se pueden especificar tipos de datos, listas de componentes y restricciones similares a las del diccionario de datos típico de una base de datos.
<b>3</b> <i>RDF + rdfschema</i>	<i>Resource Description Framework</i> (RDF) es un modelo de representación de metadatos que, entre otras cosas, permite representar recursos digitales tales como sitios o páginas web. RDF está concebido para representar cualquier clase de recursos (no solamente páginas publicadas en la web). <i>RDF Schema</i> , por su parte, es una extensión de RDF que aporta un lenguaje con mayor capacidad para representar relaciones semánticas complejas.
<b>4</b> <i>Ontology vocabulary</i>	Una <i>ontología</i> es una especificación formal de un dominio del conocimiento que, en su expresión más simple, se identifica con una taxonomía. Una taxonomía consiste en una jerarquía de conceptos y sus relaciones del tipo clase-subclase. Una ontología formaliza la relación de clase, añade otras relaciones y especifica propiedades para individuos y clases. <i>Ontology-vocabulary</i> se refiere a una ontología concreta sobre un dominio concreto del conocimiento.
<b>5</b> <i>Logic</i>	En este contexto, <i>logic</i> se refiere al estudio de las reglas formales que permiten determinar si un razonamiento se sigue de sus premisas. La lógica estudia, por tanto, la estructura de los razonamientos válidos. Se espera que los ordenadores del futuro puedan efectuar razonamientos sobre los recursos y servicios de la Web combinando los conocimientos expresados en las ontologías, los hechos declarados en los metadatos y la aplicación de reglas lógicas.
<b>6</b> <i>Proof</i>	En este contexto, <i>Proof</i> (prueba) significa demostración [matemática]. Se considera que un ordenador alcanza la máxima fiabilidad en sus razonamientos cuando es capaz de realizar demostraciones o, lo que es lo mismo a efectos prácticos, cuando es capaz de justificar el motivo por el cual tomó (o aconsejó tomar) una decisión.
<b>7</b> <i>Trust (+ Digital Signature)</i>	La última capa, <i>Trust</i> (confianza) debe servir para otorgar confianza a las transacciones en la Web a través que se llevarán a cabo no solamente entre usuarios y sitios web sino también entre programas de software; y todo ello tanto en el plano C2B ( <i>consumer to business</i> ) como en el B2B ( <i>business to business</i> ). La <i>Digital Signature</i> (firma digital) proporcionará soporte específico a esta capa, tal como muestra el diagrama.

Como hemos señalado, hasta ahora, solamente las tres primeras capas (las filas sombreadas) disponen de algún nivel de desarrollado, y ni siquiera lo están en su totalidad. Respecto a las

siguientes capas, la verdad es que se ha generado una gran cantidad de literatura técnica, científica y filosófica, pero a todos los efectos prácticos están totalmente en pañales.

Por lo tanto, en estos momentos, y al menos durante los próximos ¿tres, cuatro años?, XML, RDF y OWL (y por este mismo orden) serán con mucha diferencia los pilares más importantes del proyecto. Sucede que, además, son los elementos de mayor importancia para la Biblioteconomía-Documentación, razón por la cual, a partir de ahora, nos centraremos en ellos.

## 5.1 XML

XML es sin ninguna duda el elemento de la Web semántica que mayor repercusión tiene ya (y que sin duda continuará teniendo en el futuro) en la Biblioteconomía-Documentación. XML es un estándar (una *recommendation* en palabras del W3C) que, junto con su norma asociada, XML Schema, permite definir tipos de documentos y los conjuntos de etiquetas necesarias para codificar tales tipos de documentos. La idea es que, una vez los documentos están marcados o codificados con un conjunto de etiquetas XML es posible procesarlos y explotarlos de forma automática con diversos propósitos, de la misma manera que un conjunto de registros de una base de datos se puede explotar de formas diversas, e incluso exportarse a diversos sistemas de gestión de bases de datos si la estructura de registros sigue algún tipo de estándar.

Uno de estos propósitos puede ser codificar los documentos una sola vez, pero poder mostrarlos a través de distintos dispositivos: un navegador de Internet como Mozilla o Explorer, la pantalla de un PDA, la pantalla de un móvil, etc., siempre con el mismo conjunto de etiquetas y simplemente aplicando cada vez una hoja de estilo distinta. Otros propósitos pueden ir desde la minería de datos (si se dispone de un conjunto de documentos suficiente) hasta la recuperación de información.

XML es, por tanto, un meta lenguaje de marcado que, por un lado proporciona la posibilidad de codificar páginas de un modo directo con etiquetas *ad-hoc* y, por otro, proporciona la posibilidad de definir esquemas y tipos de documentos que, a su vez, permiten crear instancias de documentos cuya adecuación puede ser validada de forma automática con programas informáticos.

Con XML se pueden diseñar lenguajes de marcado muy estructurados y muy explícitos en los cuales, en lugar de etiquetas como <b> e <i>, podemos disponer de etiquetas como <título>, <subtítulo>, <capítulo>, <autor>, <institución>, <ciudad>, etc. De este modo, si una empresa o institución necesita almacenar y procesar información sobre los curriculum vitae de sus empleados, puede desarrollar un Schema XML que le permita crear documentos XML bien formados que dispongan de etiquetas como <lugar\_nacimiento>, <titulacion\_academica>, <idiomas\_hablados>, <experiencia\_laboral>, etc.

XML puede ser visto, a la vez, como un formato de publicación que permite crear lo que algunos autores denominan *text-centric documents*, o puede ser visto como un sistema de intercambio y procesamiento de datos que permite crear en este caso lo que se denomina *data-centric documents*.

Como sistema de edición y publicación es un formato que utilizan ya numerosas aplicaciones ofimáticas. La más importante en el mundo del open software es OpenOffice. Las dos siguientes ilustraciones muestran una captura de un texto editado con OpenOffice y el código fuente del mismo texto generado en formato XML de forma automática por OpenOffice:

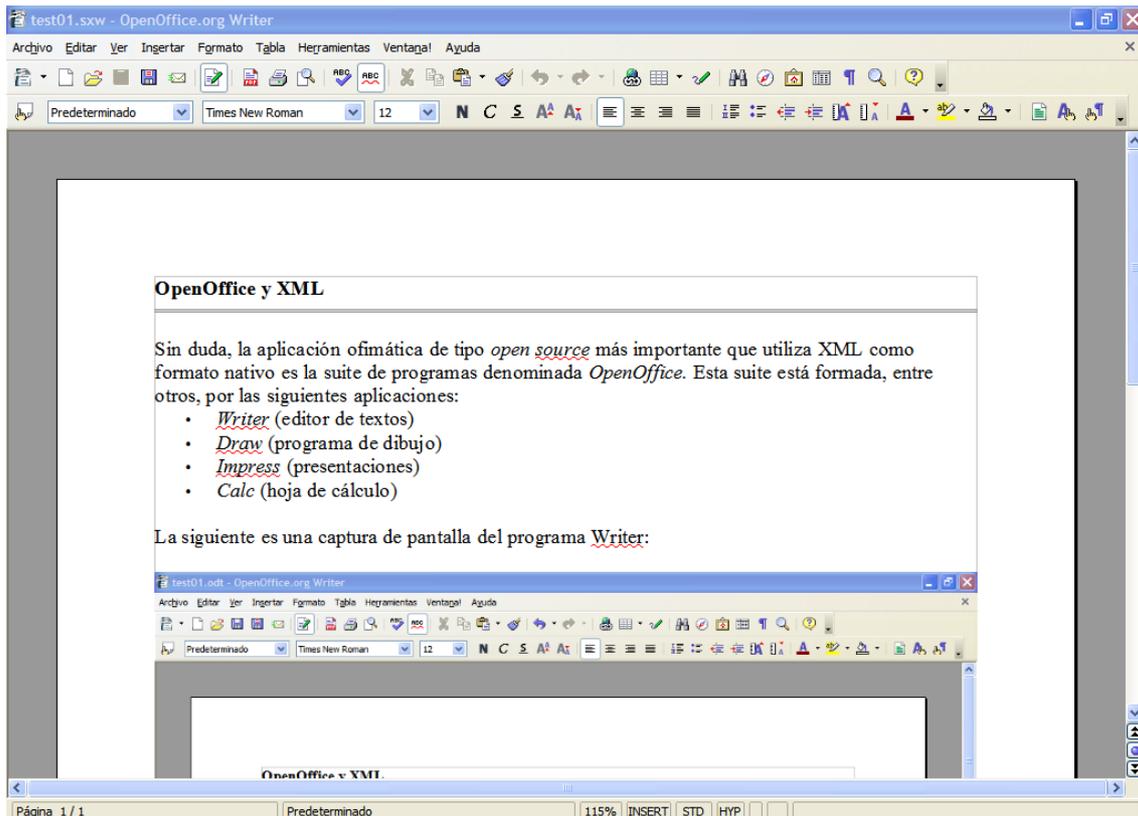


Figura 3: OpenOffice utiliza XML como formato nativo de manera transparente al usuario

```

<text:p text:style-name="P1">OpenOffice y XML</text:p>
<text:p text:style-name="Horizontal Line" />
<text:p text:style-name="Standard">
Sin duda, la aplicación ofimática de tipo <text:span text:style-name="T1">open
source</text:span> más importante que utiliza XML como formato nativo es la suite de
programas denominada <text:span text:style-name="T1">OpenOffice.</text:span>
<text:span text:style-name="T2">Esta suite está formada, entre otros, por las siguientes
aplicaciones:</text:span>
</text:p>
<text:ordered-list text:style-name="L1">
<text:list-item>
<text:p text:style-name="P2">
<text:span text:style-name="T1">Writer</text:span> (editor de textos)
</text:p>

```

Figura 4: Vista parcial del código fuente XML generado por *OpenOffice* del documento de la imagen anterior

Algunas observaciones sobre lo anterior: en primer lugar, el uso que hace OpenOffice de XML es transparente al usuario, es decir, éste no necesita saber nada sobre XML y, de hecho, OpenOffice no abruma al usuario informándole siquiera sobre este formato salvo que el usuario desee bucear en las interioridades técnicas del programa. En segundo lugar, OpenOffice permite guardar documentos en distintos lenguajes o aplicaciones XML, entre ellos, el formato DocBook para codificar libros y documentos técnicos y el formato XML de las aplicaciones de Microsoft Office. Por último, como podemos ver un documento codificado con XML guarda enormes semejanzas con el código fuente HTML de una página web.

El motivo es que ambos lenguajes, HTML y XML, derivan del mismo metalenguaje, denominado SGML<sup>4</sup> y, por tanto, ambos comparten la misma filosofía básica y la forma de actuación idéntica: los documentos consisten en texto rodeado por parejas de marcas específicas (como las que conocemos en el caso de HTML: <h1>... </h1>, <p>... </p>), con una de ellas actuando como inicio de la marca (<h1>) y otra como final de la marca (</h1>). El texto contenido entre una marca de inicio y una marca de final es el contenido afectado por dichas marcas. Estas marcas se denominan etiquetas y el conjunto de las etiquetas de inicio y final más el texto afectado se denomina elemento.

Miller y Clarke (2004) examinan las características que hacen de XML un estándar especialmente importante para el mundo de la Biblioteconomía-Documentación y las agrupan en tres ejes principales:

1. *No propiedad / Interoperabilidad / Neutralidad de plataforma*
2. *Longevidad / Persistencia / Utilización futura*
3. *Separación de contenido, visualización y funcionalidad / Reutilización*

El primer eje (*No propiedad*, etc.) se refiere al hecho de que XML es un sistema abierto, no propietario e independiente de la plataforma, es decir, independiente tanto de sistemas operativos (Windows, Unix, etc.) como de ordenadores concretos (PC, Macintosh, Sun, etc.). Esto significa que puede funcionar exactamente igual de bien en todos ellos.

El segundo (*Longevidad*, etc.) incide en los aspectos, tan propios del mundo bibliotecario y archivístico, de la preservación de la información. En realidad, esta faceta de XML deriva directamente de la primera: puesto que la información codificada en XML es independiente de la plataforma y el código fuente puede interpretarse directamente por el ojo humano, su persistencia y conservación están garantizadas para el futuro.

Por último, el tercer eje (*Separación de contenido*, etc.) incide en uno de los aspectos más interesantes e inteligentes de XML: al insistir en separar el marcado del contenido de su presentación, XML es un instrumento de primer orden para procesar información. En efecto, la información pueden almacenarse en un único formato y reutilizarse en tantos formatos como sea necesario: HTML, PDF, página web, documento impreso, monitor de ordenador, PDA, etc.

Lo cierto es que si cada uno de los tres ejes mencionados justificaría por sí solo el uso de XML en el mundo de la biblioteconomía-documentación, los tres juntos deberían formar un argumento imbatible.

Por el momento, entre las aplicaciones de XML a la biblioteconomía-documentación podemos mencionar sendos proyectos liderados por la *Library of Congress* para migrar el formato MARC al entorno XML ([www.loc.gov/standards/marcxml/](http://www.loc.gov/standards/marcxml/)) y de hacer lo mismo con el formato de codificación de archivos EAD (Encoded Archival Description).

La inmensa mayoría de los proyectos actuales de bibliotecas digitales que se están desarrollando en diversos ámbitos temáticos y geográficos se apoyan, en más de una forma, en el uso de XML, bien sea por la forma de codificación de la información, con formatos como TEI o DocBook o bien sea por la forma de codificación de los metadatos de cada documentos, con formatos como Dublín Core expresado en XML/RDF.

Por último, podemos señalar el movimiento *Open Access* como otra de las formas más interesantes en las cuales XML tiene incidencia en el mundo de la documentación, en este caso en el mundo de la información y documentación científica a través de los repositorios creados

---

<sup>4</sup> SGML son las siglas de *Standard Generalize Markup Language*, y fue desarrollado originalmente en la empresa IBM en la década de los 70. Actualmente, y desde hace años, es una norma ISO para generar etiquetas de marcado de documentos. De esta norma proceden los lenguajes HTML, XHTML y XML.

por diversas instituciones, particularmente, universidades de artículos, informes y resultados de investigación gracias al movimiento *Open Archives Initiatives* ([www.openarchives.org/](http://www.openarchives.org/)).

## 5.2 Metadatos y RDF

El segundo gran componente de la Web semántica son los metadatos, la tercera capa en el diagrama de Berners-Lee. Los documentalistas están familiarizados con los metadatos desde hace décadas, aunque con otro nombre. Los metadatos son información sobre la información (o datos sobre datos) y como tal son, en realidad, una antigua fórmula. Los catálogos de las bibliotecas son una clase de metadatos en tanto informan sobre los documentos disponibles en la biblioteca. Lo mismo sucede con las referencias bibliográficas de una base de datos documental como LISA (<http://www.csa.com/factsheets/lisa-set-c.php>).

Base de datos LISA: Library and Information Science Abstracts

**Título** Attention web designers: You have 50 milliseconds to make a good first impression!

**Autor** Lindgaard, Gitte; Fernandes, Gary; Dudek, Cathy; Brown, J

**Fuente** Behaviour and Information Technology; 25 (2) Mar 2006-Apr 2006, pp.115-126

**ISSN** 0144-929X

**Descriptor**  World Wide Web  Web sites  Design  Usability  Evaluation  Performance measures  Usability testing

**Nueva búsqueda utilizando los términos marcados:**  Utilice AND para restringir la búsqueda  Utilice OR para ampliar la búsqueda

**Resumen** Three studies were conducted to ascertain how quickly people form an opinion about web page visual appeal. In the first study, participants twice rated the visual appeal of web homepages presented for 500 ms each. The second study replicated the first, but participants also rated each web page on seven specific design dimensions. Visual appeal was found to be closely related to most of these. Study 3 again replicated the 500 ms condition as well as adding a 50 ms condition using the same stimuli to determine whether the first impression may be interpreted as a 'mere exposure effect' (Zajonc 1980). Throughout, visual appeal ratings were highly correlated from one phase to the next as were the correlations between the 50 ms and

Figura 5: Un típico registro bibliográfico consiste en un conjunto de metadatos (señalados con flechas). En este caso, vemos los metadatos correspondiente a un artículo científico (tomado de la base de datos LISA)

En el lenguaje documental clásico se habla de documentos secundarios y de documentos primarios mientras que la terminología propia de la web habla de recursos y de metadatos. En ambos casos estamos delante de información sobre la información (o de datos sobre datos). La tabla siguiente ilustra esta relación.

**Tabla 3: Relación entre la terminología documental y la Web**

Terminología "clásica"	Terminología de la Web	Significado común
Documento secundario	Metadatos	Datos sobre datos (o información sobre información)
Documento primario	Recurso	Información original

Ahora bien, la Web semántica ha puesto de manifiesto que, para diversos colectivos existen diversos tipos de metadatos. Por ejemplo, para algunos documentalistas los metadatos son, sobre todo, los datos que ayudan a la recuperación de información (como las palabras clave que se usan para indizar un documento) pero, para archiveros y documentalistas los metadatos son sobre todo descriptivos (como la fuente, la autoría o la fecha de publicación). Además de estas diferencias de enfoque (que no obstante se complementan entre sí) existen diferencias de sector.

Por ejemplo, un archivo de imágenes de arte necesita un conjunto de metadatos distinto del que necesita un repositorio de tesis doctorales. Las diferentes funciones que cumple un documento añaden nuevas necesidades: según el contexto, se requieren metadatos sobre propiedad intelectual, sobre necesidades técnicas de reproducción o sobre condiciones de conservación.

Además, el enfoque propio de la Web coincide también con una cierta tradición documental consistente en entender el alcance de los metadatos hasta el concepto de recurso. De acuerdo con esta idea, un recurso no solamente puede ser un documento, sino que las personas o los objetos materiales también son recursos<sup>5</sup>. Por tanto, se requieren conjuntos de metadatos también para otros tipos de entidades como los mencionados.

En la Web el sistema que permite utilizar metadatos para describir recursos (típicamente sitios web) recibe el nombre de *Resource Description Framework* (RDF). Ahora bien, el proyecto de la Web semántica, en lugar de intentar proponer conjuntos concretos de metadatos para todos los propósitos y contextos imaginables (misión que, sin duda, acabaría en fracaso), propone un sistema abstracto de validez universal que debe servir para expresar cualquier conjunto, presente o futuro, de metadatos.

Como en otros elementos de la Web semántica, la base conceptual de RDF proviene de la lógica, pero en este caso con componentes de la lingüística. Todo el sistema RDF parte de tres entidades lógicas:

1. Recursos
2. Propiedades
3. Valores

Que corresponde a los elementos de la lingüística:

1. Sujeto
2. Predicado
3. Objeto

Con los tres elementos anteriores podemos formar declaraciones (*statements*) sobre recursos del tipo: el recurso *X* tiene la propiedad *Y* con el valor *P*. La tabla siguiente expresa las equivalencias de los componentes básicos de RDF:

**Tabla 4: Equivalencia lógico-lingüísticas en una declaración RDF**

Declaración			
Término Lingüístico	Sujeto	Predicado	Objeto
Término Lógico	Recurso	Propiedad	Valor
<b>Ejemplo</b>	hpDeskjet9800	tipoImpresión	Inyección térmica de tinta

Los recursos pueden ser sitios o páginas web, pero también pueden ser cosas que no están en la web, como personas o cualquier objeto del mundo real o conceptual. Las propiedades son las características relevantes de los recursos. Por ejemplo, en relación a las páginas web, el autor y el idioma son dos características relevantes de la misma en casi cualquier contexto y, por tanto, son dos de los atributos típicos a la hora de representar sitios web. Por último, los valores son los datos en los que se concreta un atributo determinado de un recurso determinado. La tabla siguiente expresa las ideas anteriores con dos ejemplos concretos aplicados a la descripción de dos sitios web utilizando Dublín Core (dublincore.org):

<sup>5</sup> Un recurso es alguna cosa que nos permite alcanzar un objetivo. En general, la información es reconocida como un recurso porque permite alcanzar objetivos de forma eficiente.

**Tabla 5: Declaracion RDF sobre dos sitios web**

<i>Recurso</i>	<i>Propiedad</i>	<i>Valor</i>
<a href="http://www.imdb.com">http://www.imdb.com</a>	dc.title	Internet Movie Database
<a href="http://allmovie.com">http://allmovie.com</a>	dc.title	All Movie Guide

De acuerdo con la tabla anterior, hemos representado dos recursos (en este caso, dos bases de datos cinematográficas) mediante una de sus propiedades, en este caso el título de la página web. Naturalmente, una representación profesional de esas bases de datos requeriría otras propiedades, pero siempre podríamos utilizar esta misma estructura, denominada triple, para expresarlos, cualquiera que fueran. Veamos, por ejemplo, la forma en que Artifact presenta metadatos sobre AllMovie.

The image shows a screenshot of the 'Artifact' website. The header includes the 'artifact' logo, a tagline 'Best of the Web for the Arts and Creative Industries', and a 'a service of the R-D-N' logo. A navigation menu contains 'Home', 'About', 'New', 'Suggest', and 'Feedback'. On the left is a vertical sidebar with categories like 'Artist's Index', 'Limelight', 'Subjects A-Z', 'Subjects List', 'Architecture', 'Communications, Media and Culture', 'Design', 'Fashion and Beauty', 'Music and the Performing Arts', 'Visual Arts', 'Advice and Guidance', 'Collections and Exhibitions', 'Image Banks', 'Teaching, Learning and Research', and 'Virtual Training'. The main content area is titled 'Back' and contains the following metadata for 'All movie guide':

- Title:** All movie guide
- URL:** <http://www.allmovie.com/> (circled in blue)
- Description:** The All Movie Guide is a database produced by the All Media Guide (AMG). This resource can be searched by film, person or keyword. The 'film finder' provides some advanced search features, such as searching for original works on which films have been based, or searching for specific film locations and technical criteria. Additional features include a "quick browser", organised by movie genre, country or time period; a glossary of movie and film-related terms; and a series of essays. The movie guide home page links to new movies, actors born today and showcases specific genres or periods. Information provided for films includes full credits, cast list, synopsis, reviews, similar films and awards. All people involved in movies within the databases are indexed, with filmographies and biographical details provided for many.
- Author:** All Media Guide
- Keywords:** AAT\*: film (performing arts); motion pictures (visual works); filmmaking; filmmakers; feature films; filmographies; producers; directors (performing arts); actors; biographies; awards; motion picture genres
- Other:** cinema (performing arts); moving pictures; movies; movie awards; film awards; movie locations; film locations; movie genres; film genres; motion picture locations
- Subject Headings:** film reference and resources
- Language(s) available on Website:** English

Annotations include a blue speech bubble labeled 'Recurso' pointing to the URL, and several red arrows pointing to the 'Description', 'Author', 'Keywords', 'Subject Headings', and 'Language(s)' fields.

Figura 6: Descripción de un sitio web. El recurso está rodeado por un círculo. Las flechas señalan los distintos metadatos y sus valores concretos (tomado de la web Artifact).

La figura anterior, tomada de Artifact ([www.artifact.ac.uk](http://www.artifact.ac.uk)) muestra una serie de metadatos sobre AllMovie. Si los observamos bien, podemos ver que pueden ser expresados según el modelo RDF sin ningún problema:

**Tabla 6: Metadatos de Artifact sobre AllMovie expresados como triples**

<i>Recurso</i>	<i>Propiedad</i>	<i>Valor</i>
<a href="http://allmovie.com">http://allmovie.com</a>	Title	All movie guide
<a href="http://allmovie.com">http://allmovie.com</a>	Author	All Media Guide
<a href="http://allmovie.com">http://allmovie.com</a>	Description	The All Movie Guide is a database produced by the All Media Guide (AMG). This resource can be searched by film, person or keyword (...)
<a href="http://allmovie.com">http://allmovie.com</a>	Keywords	Film (performing arts); motion pictures (visual works); filmmaking; (...)
<a href="http://allmovie.com">http://allmovie.com</a>	Subject Heading	Films reference and resources
<a href="http://allmovie.com">http://allmovie.com</a>	Language	English

En realidad, los seres humanos se supone que estamos absolutamente familiarizados con esta forma de describir las cosas. Si decimos el “coche cuyo color es rojo”, o “el libro cuyo autor es Umberto Eco” estamos usando la misma estructura que RDF. Lo que sucede es que, en realidad, nadie habla así (perdería a todos sus amigos rápidamente). Más bien nos referimos simplemente “al coche rojo” o “al libro de Umberto Eco” con lo cual la estructura de triples pasa desapercibida porque no incluimos el nombre del atributo, sino directamente su valor. La cuestión es que los ordenadores no pueden procesar de esta forma la información y debe hacerse explícito el nombre del atributo.

Originalmente, la representación de RDF fue concebida en forma de grafos (un sistema de diagramas de inspiración matemática<sup>6</sup>), de manera que los recursos tenían forma de elipse, el atributo en forma de flecha y el valor del atributo en forma de rectángulo. De este modo, el título de un sitio empleando la norma Dublín Core se representaría mediante el siguiente grafo:

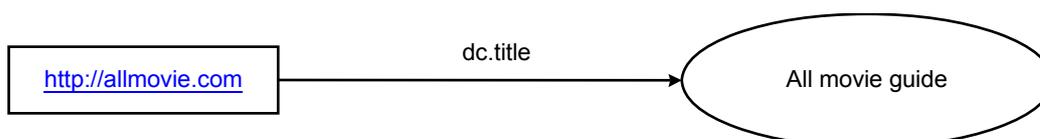


Figura 7: Un sitio web representado como un grafo

En el sistema gráfico empleado por RDF el recurso se representa, como puede verse, mediante un cuadrado que se identifica mediante la URI correspondiente (en este caso, un URL). El nombre de la propiedad (dc.title) también necesita un URI que, en este caso hemos representado simplemente por dc.title para simplificar (el URI completo sería otro URL tal como “<http://purl.org/dc/elements/1.1/dc.title>”). Finalmente, el valor de la propiedad es un literal, es decir, un texto o un cadena de caracteres (“All movie guide”).

Naturalmente, no es muy práctico describir miles de recursos, cada uno de ellos con decenas de atributos mediante grafos. Por este motivo, pronto se sugirieron formas de expresar estos grafos virtuales mediante una codificación en base a líneas de texto, operación que recibe el nombre de serialización. Por ejemplo, la tabla siguiente muestra la misma información que el grafo anterior, pero serializada (en este caso, usando RDF/XML y Dublín Core):

<sup>6</sup> Un grafo es la representación de un objeto, de una relación o de una estructura mediante puntos (nodos) conectados por líneas (arcos). En matemáticas, la teoría de grafos sirve para representar problemas y sus posibles soluciones. En otras áreas de conocimiento, como en el caso de la Web semántica, los grafos se utilizan para expresar relaciones.

```

<?xml version="1.0" ?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc="http://purl.org/dc/elements/1.1/">
<rdf:Description rdf:about="http://allmovie.com">
<dc:title>All movie guide</dc:title>
</rdf:Description>
</rdf:RDF>

```

Figura 8: Serialización del grafo de la figura 7

Veamos un segundo ejemplo, ahora con varias propiedades de un mismo recurso. En este caso: el título, la descripción, la fecha, el formato, el lenguaje y la autoría (*contributor*). En este caso, se trata de los metadatos del sitio web de *Dublin Core Metadata Initiative* ([dublincore.org/](http://dublincore.org/)) serializados también mediante XML/RDF:

```

<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc="http://purl.org/dc/elements/1.1/">
<rdf:Description rdf:about="http://dublincore.org/">
<dc:title>Dublin Core Metadata Initiative (DCMI) Home Page</dc:title>
<dc:description>The Dublin Core Metadata Initiative is an open forum engaged
in the development of interoperable online metadata standards that support a
broad range of purposes and business models. DCMI's activities include
consensus-driven working groups, global conferences and workshops, standards
liaison, and educational efforts to promote widespread acceptance of metadata
standards and practices.</dc:description>
<dc:date>2006-03-13</dc:date>
<dc:format>text/html</dc:format>
<dc:language>en</dc:language>
<dc:contributor>Dublin Core Metadata Initiative</dc:contributor>
</rdf:Description>
</rdf:RDF>

```

Figura 9: Metadatos tomados de la página de Dublin Core Metadata Initiative ([dublincore.org/index.shtml](http://dublincore.org/index.shtml))

Dado que en un capítulo de este mismo libro se trata de forma amplia RDF y los programas que permiten editar archivos RDF damos por concluida aquí los aspectos más técnicos de este formato.

Sin embargo, lo que nos gustaría destacar es la enorme semejanza de un conjunto dado de metadatos RDF con la tabla de una base de datos relacional. En primer, podemos observar que la tabla n. X se puede transformar en una tabla de una base de datos relacional simplemente cambiando el título de las columnas.

**Tabla 7: Metadatos de Artifact sobre AllMovie expresados como una tabla relacional**

<b>ID</b>	<b>Title</b>	<b>Author</b>	<b>Description</b>	<b>Keywords</b>
<a href="http://allmovie.com">http://allmovie.com</a>	All movie Guide	All Media Guide	The All Movie Guide is a database produced by the All Media Guide (AMG). This resource can be searched by film, person or keyword (...)	Film (performing arts); motion pictures (visual works); filmmaking; (...)

En efecto, en el modelo relacional, cada entidad es una fila de una tabla, y cada columna es un atributo de la entidad. En la siguiente tabla podemos ver las equivalencias entre ambos modelos.

**Tabla 8: Equivalencias entre modelo relacional y RDF**

<i>Modelo relacional</i>	<i>Modelo RDF</i>	<i>Significado común</i>
Entidad	Recurso	Aquella parte del mundo real o conceptual sobre la cual se aporta alguna (meta) información
Atributo	Propiedad	Características relevantes de la entidad o del recurso
Valor	Valor	Dato concreto que asume el valor de una entidad o de un recurso

Cabe señalar un aspecto muy importante sobre RDF que no siempre es bien entendido. Aunque la inspiración inicial nació en relación a los metadatos y la descripción de recursos tal como la entendemos en las Ciencias de la Documentación, lo cierto es que RDF tiene aplicaciones muy diversas.

En primer lugar, la imagen que tenemos los documentalistas sobre los metadatos es la que responde al conocido esquema *documento secundario > documento primario*, donde el documento secundario consiste en un conjunto de pares propiedad/valor (conjunto de propiedades y sus valores correspondientes) que describen al documento original y la relación consiste en que el documento secundario *describe* al primario.

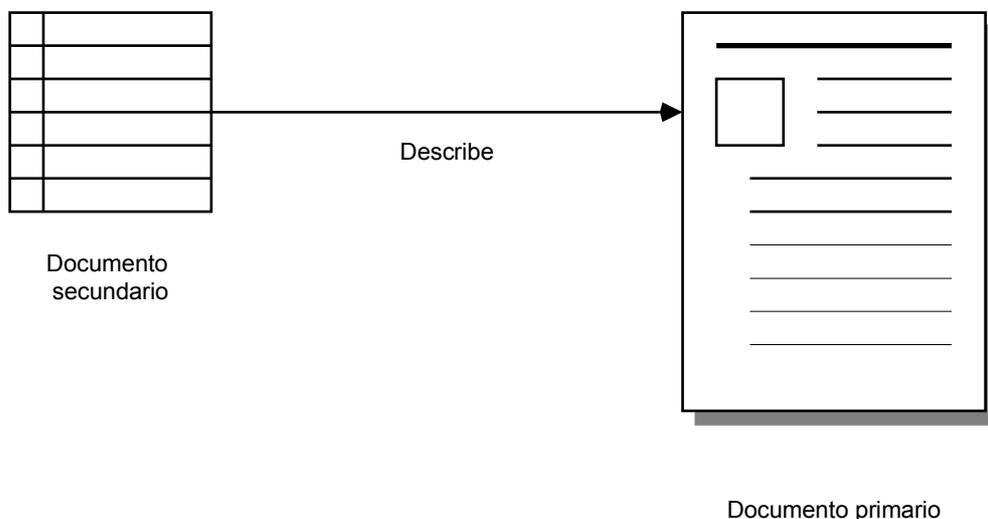


Figura 10: Los metadatos tal como se conciben en Documentación: un documento primario (p.e. un artículo de revista) es descrito mediante un documento secundario (p.e., mediante el correspondiente registro de una base de datos bibliográfica)

En segundo lugar, para los documentalistas, aunque lo siguiente sea muy explícito, cabe hacer hincapié en ello: los metadatos *siempre* describen documentos, y si hablamos de la Web, entonces los metadatos siempre describen páginas (o sitios) web.

Sin embargo, para comprender el verdadero alcance de RDF cabe señalar que este modelo ambas presuposiciones quedan desbordadas ampliamente: por un lado, el concepto de metadato en la Web semántica no se restringe al concepto de la descripción (o representación) de algo. Para la Web semántica, una determinada estructura de información es una clase de metadatos.

Por ejemplo, la estructura concreta de la información mediante la cual las compañías de aviación registran información sobre sus vuelos, rutas, asientos ocupados, asientos disponibles, etc., son metadatos. En este caso, los metadatos son la estructura de pares atributo/valor que sirven para registrar una actividad del mundo real: datos sobre vuelos, reservas de asientos, etc. Este espíritu, procedente del mundo de la ingeniería del software, está presente en RDF y es el que explica su semejanza con el mundo de las bases de datos relacionales.

Por otro lado, para el modelo RDF las clases de recursos que se pueden describir no se agotan en las páginas web. Todo lo contrario: en RDF cualquier cosa que se puede nombrar o señalar de manera específica (esto es, cualquier cosa que tenga una forma inequívoca de ser identificada) es un recurso. Por tanto, los libros impresos son recursos para RDF desde el momento que tienen un ISBN, las personas son recursos desde el momento en que tienen un número de DNI (o de pasaporte, o de seguridad social), las empresas son recursos por el mismo motivo (tienen número de identificación fiscal), etc.

En general, cualquier cosa sobre la que se pueda decir cualquier cosa es susceptible de ser tratada en RDF. En último extremo, si no disponemos de una identificación previa para “la cosa”, podemos asignar nosotros una identificación arbitraria con la única condición de que sea única.

Ignoramos cómo evolucionará la aplicación de estos conceptos de RDF a la hora de su aplicación al mundo práctico. Sin embargo, podemos decir que, por el momento, no parecen afectar al trabajo característico de la Documentación. El motivo por el que los señalamos es para ayudar a entender algunas características de RDF que, de otro modo, parecen innecesariamente crípticas.

Afortunadamente, esta amplitud de miras de RDF no es un obstáculo para su aplicación al mundo de la documentación, sino todo lo contrario: una de las más importantes y significativas aplicaciones de RDF consiste en la descripción de recursos digitales utilizando la norma Dublin Core. Norma que, como es sabido, consiste precisamente en aplicar la filosofía documental indicada antes, a saber, los metadatos Dublin Core son descripciones al más puro estilo bibliográfico (título, autor, fecha de publicación, palabras clave, etc.), es decir documentos secundarios aplicados a sitios web, es decir documentos primarios.

### 5.3 Ontologías

El término ontología ha gozado de varias vidas. Originalmente, el término provenía de la Filosofía clásica. En ese contexto, la ontología era una parte de la metafísica que se ocupaba de estudiar la naturaleza de la existencia. Era una especialidad que gozó del favor de los filósofos casi desde el nacimiento mismo de la Filosofía en la Grecia clásica (aunque el término “ontología” aparece en el siglo XVII). La llegada de nuevas corrientes filosóficas a finales del siglo XIX de corte anti metafísico hicieron que el término perdiera gran parte de su vigencia fuera de algunas escuelas de pensamiento minoritarias o muy especializadas.

Sin embargo, mucho después, tanto el colectivo de los informáticos como el colectivo de los lingüistas rescataron el término para darle significados distintos e incorporarlo al lenguaje de la ciencia.

En particular, los estudios de Inteligencia Artificial recuperaron el término para designar esquemas conceptuales formalizados sobre algún aspecto de la realidad, con la finalidad de facilitar su reutilización en diferentes contextos o la comunicación entre diferentes sistemas, casi siempre con el telón de fondo de la construcción de sistemas expertos.

En el campo de la terminología apareció también el uso del término ontología. En este caso para referirse a una clase de compilaciones léxicas dónde además de indicar la clase de cada término (verbo, nombre, etc.) se indican también las relaciones de sinonimia y se establecen árboles conceptuales que presentan las relaciones jerárquicas y de inclusión entre términos (hipónimos, hiperónimos, homónimos, merónimos, etc.). En este caso, el objetivo de las ontologías para los lingüistas solía ser la creación de diccionarios y enciclopedias.

Por último, aparecieron usos ocasionales del término en ámbitos dispares. Por ejemplo, los responsables de Yahoo denominaron ontología a su sistema de clasificación. No es extraño por tanto que también se diera un uso incipiente (y bastante dubitativo) en Ciencias de la Documentación para referirse a algunos lenguajes documentales. Por ejemplo, en algunas ocasiones, MESH, el tesoro de medicina de la Biblioteca Nacional de Medicina de EE.UU.) ha sido presentado como una ontología.

En los últimos años, el proyecto de la Web semántica ha servido para asentar lo que podríamos denominar el uso “moderno” del término ontología. Este uso actual, como no podía ser de otra forma, está muy cercano al que tenía el concepto en el campo de la Inteligencia Artificial, aunque ahora lo observamos bajo una perspectiva nueva, a saber, la de la Web semántica.

En este nuevo contexto, una de las definiciones más citada es la debida a Gruber (1993) según la cual una ontología es “la especificación de una conceptualización”. Ahora bien, para entender esta definición tan afortunada, pero a la vez tan compacta que dificulta su entendimiento, vale la pena acudir a la fuente original (muy citada pero probablemente muy poco consultada):

Un cuerpo de conocimiento formalmente representado se basa en una conceptualización: los objetos, conceptos y otras entidades cuya existencia se presume en área de interés, así como las relaciones que mantienen entre ellas. (Gruber, 1993: 2).

Es decir, para representar un cuerpo de conocimiento primero debemos conceptualizarlo mediante la especificación de las entidades que forman parte de ese cuerpo, así como debemos especificar las relaciones que tales entidades mantienen entre ellas. A continuación, Gruber indica:

Una conceptualización es una abstracción, una visión simplificada del mundo que queremos representar para algún propósito. Cada base de conocimiento, cada sistema

basado en conocimiento, o cada agente de conocimiento está comprometido con alguna conceptualización, implícita o explícita (Gruber, 1993: 2).

Es a partir de aquí cuando Gruber presenta su conocida definición (y cuando adquiere sentido) según la cual una ontología “es la especificación de una conceptualización”. El gráfico siguiente presenta un ejemplo de clases y subclases de una ontología sobre periféricos de ordenador.

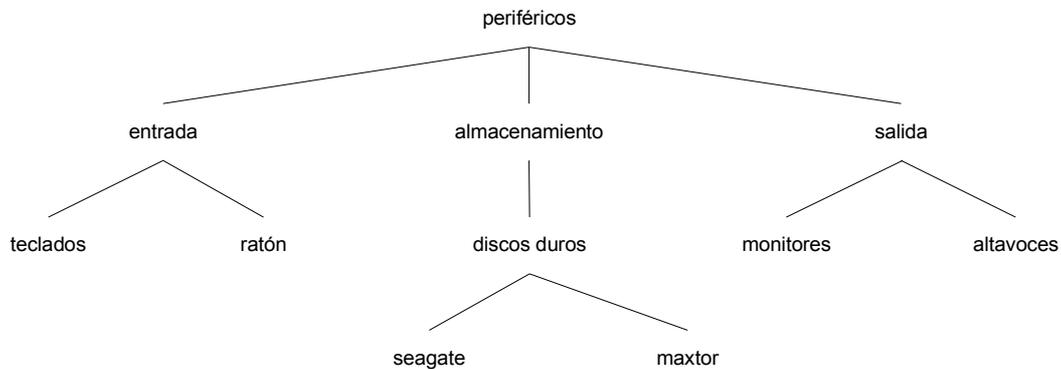


Figura 11: Clases y subclases de una ontología sobre periféricos de ordenador

No obstante, aunque la definición de Gruber ha resultado ser profética no debemos perder de vista que la definición anterior es previa, no al proyecto de la Web semántica, sino a la Web a secas. Actualmente, en el contexto de la Web semántica, una ontología presenta las siguientes características:

- Se refiere siempre a un dominio concreto del conocimiento. De hecho, se han desarrollado ontologías para sectores tan diversos como las pizzas y las ciencias de la vida; el derecho de autor y la astronomía, los medios de comunicación y el periodismo, etc. El sitio web del proyecto DAML para el desarrollo de ontologías ofrece un directorio de ontologías ([www.daml.org/ontologies/](http://www.daml.org/ontologies/)). A inicios del 2006, presentaba casi trescientas.
- En segundo lugar, los componentes de la ontología y sus relaciones entre ellos están concebidos en clave de lógica formal utilizando los conceptos de *clase* (por ejemplo, la clase de los “discos duros”), *propiedades* (por ejemplo, “capacidad de almacenamiento”) e *individuos* (por ejemplo, el disco duro marca “*Maxtor DiamondMax11*”). Además, una ontología describe las siguientes relaciones entre los componentes: individuos con clases, o sea la relación “x es una instancia de y” (por ejemplo: “*Maxtor DiamondMax11* es un disco duro”); individuos con propiedades, o sea, la relación “x tiene el valor y para la propiedad p” (por ejemplo, *DiamondMax11* tiene el valor de “500GB” para la propiedad “capacidad de almacenamiento”); y clases con propiedades, o es decir, relaciones de restricción, por ejemplo, los discos duros no son separables de la unidad central (a diferencia por, ejemplo, de las unidades de memoria USB).
- En tercer lugar, a diferencia de la idea que imperaba en la IA “clásica”, la ontología debe estar codificada en un lenguaje compatible con el entorno abierto y descentralizado característico de la Web. Esto implica un doble requerimiento: por un lado que la ontología esté codificada en el lenguaje propio de la Web semántica (OWL)

para que pueda ser (re)utilizada a través de diversos sistemas, en diversos contextos y por parte de agentes de usuario heterogéneos. Por otro lado, implica que las ontologías estarán publicadas y disponibles en ordenadores (servidores) interconectados en la Web.

## 5.4 OWL

*OWL Web Ontology Language* (OWL) es el lenguaje estándar de la Web semántica para expresar y codificar ontologías. De acuerdo con el W3C:

OWL está concebido para ser utilizado cuando la información contenida en los documentos necesita ser procesada por aplicaciones informáticas, en oposición a las situaciones donde el contenido solamente debe ser presentado a seres humanos. OWL puede ser utilizado para representar explícitamente el significado de términos en vocabularios y las relaciones [semánticas] entre esos términos (<http://www.w3.org/TR/2004/REC-owl-features-20040210/>)

OWL permite formalizar las relaciones entre las clases aún más que RDF Schema indicando aspectos básicos para el razonamiento como la existencia de clases disjuntas. Por ejemplo: “los periféricos de salida no son periféricos de almacenamiento”, esto es, la clase de los periféricos de salida es disjunta a la clase de los periféricos de almacenamiento. También permite expresar la cardinalidad, es decir, el número de elementos que pueden componer una clase, por ejemplo, “un libro puede tener uno o varios autores” (la cardinalidad de los autores de un libro es uno o más de uno), o bien “un libro solamente puede tener un ISBN” (la cardinalidad del ISBN de los libros es exactamente uno). Puede expresar igualdad o equivalencia entre clases, características y restricciones de las mismas, etc.

OWL utiliza RDF/XML para representar y codificar las ontologías. OWL sigue la tendencia tan característica del W3C de proceder mediante “extensiones”. Por tanto, OWL es una extensión de RDF que añade elementos como los mencionados anteriormente para describir características y clases. A modo de ilustración, en la figura siguiente podemos ver parte de la ontología anterior sobre periféricos de ordenador, pero ahora representada en OWL:

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
...
<owl:Class rdf:ID="perifericos">
<rdfs:comment>
Los periféricos de ordenador están conectados a la CPU pero no forman parte de
ella
</rdfs:comment>
</owl:Class>
<owl:Class rdf:ID="entrada">
<rdfs:comment>
Los periféricos de entrada son una subclase de periféricos de ordenador.
</rdfs:comment>
<rdfs:subClassOf rdf:resource="#perifericos" />
</owl:Class>
<owl:Class rdf:ID="teclados">
<rdfs:comment>
Los teclados son una subclase de los periféricos de entrada.
</rdfs:comment>
<rdfs:subClassOf rdf:resource="#entrada" />
<rdfs:subClassOf rdf:resource="#perifericos" />
</owl:Class>
...
</rdf:RDF>
```

Figura 12: Representación en lenguaje OWL de la ontología del gráfico anterior

Podemos ver por el ejemplo anterior que un documento OWL es, ante todo un documento XML tal como se declara en la primera línea de la ontología (`<?xml version="1.0">`). A continuación, el ejemplo anterior nos muestra también que el elemento raíz de la ontología es el elemento RDF (`rdf:RDF`).

La idea esencial es que en algún momento del futuro, pero *no antes del 2010*, la Web estará poblada por un número de ontologías que permitirán a los ordenadores realizar inferencias sobre la información publicada en la Web. Estas ontologías podrán estar publicadas en los mismos servidores que sirvan para publicar las páginas web o en servidores específicos compatibles con la arquitectura de la Web.

En este sentido, cabe señalar que algunas previsiones sitúan el desarrollo de la Web semántica, incluyendo la última capa (*Proof*), no antes del 2012. Como las ontologías están situadas casi en la parte superior de este modelo de capas, es por ello que nosotros indicamos “no antes del 2010”. No obstante, por motivos que argumentaremos a continuación es posible que incluso esa fecha resulte demasiado optimista y el uso efectivo de ontologías se retrase de manera indeterminada mucho más allá de las fechas barajadas hasta el momento.

## 6 Problemas en el paraíso

El W3C se complace con el uso de un registro de lenguaje que enmascara las dificultades reales del proyecto de la Web semántica. De este discurso, por desgracia se contagian la mayor parte de los libros y artículos que se publican sobre el tema. A veces resulta lastimoso leer tanta literatura sobre la Web semántica como si ésta no solamente fuera factible a corto plazo, sino como si ya fuera una realidad cumplida.

Nosotros entendemos que, por el contrario, es importante destacar estas dificultades, en primer lugar por simple respeto a la verdad (o al menos lo que nosotros honestamente como tal), pero también por razones de eficiencia: si de verdad creemos en el proyecto, no es eficiente disimular sus dificultades.

Nadie puede discutir el acierto de Unicode, la racionalidad del sistema URI ni mucho menos el enorme logro que ha significado el lenguaje XML. También aparecen justificadas las esperanzas que ha despertado RDF como modelo de metadatos. Sin embargo, a partir de aquí, todo lo relativo a las ontologías (OWL) y el uso de la lógica por parte de los ordenadores para realizar razonamientos y demostraciones está mucho menos claro.

Incluso aspectos aparentemente inocentes del proyecto, como imaginar una Web poblada exclusivamente (o mayoritariamente) por páginas impecablemente codificadas en XHTML y con metadatos técnicamente correctos y éticamente adecuados resultan fuertemente contestados por la realidad de la Web tal como es hoy por hoy. Ciertamente, esta realidad puede cambiar en el futuro, pero para que la dimensión del cambio sea creíble para la década del 2010, deberíamos empezar a tener indicadores cada vez más solventes conforme vamos avanzando hacia esa fecha.

## 7 Migración: un diseño experimental

En este sentido, los resultados tangibles de la Web Semántica parecen darse de forma escasa y en contextos reducidos en la Web actual. Por lo que hemos señalado hasta ahora, puede verse que el inicio de la migración hacia la Web Semántica requiere, al menos estas tres condiciones:

1. *Calidad del código fuente*: código fuente de las páginas web consistente y sin errores
2. *Uso de metadatos*: páginas web descritas mediante metadatos en suficiente cantidad y calidad
3. *Uso de RDF*: codificación de los metadatos mediante RDF

Estos tres elementos aparecen como los indicadores más solventes de la evolución hacia la Web Semántica de acuerdo con la propia hoja de ruta del W3C tanto si nos atenemos a su modelo de 7 capas ya analizado, como si consideramos las características intrínsecas del proyecto

En primer lugar, parece evidente que la Web semántica no será posible con el nivel de errores en el código fuente de las páginas web que presentan actualmente. Estos errores han sido sistemáticamente compensados con la alta tolerancia de los navegadores al código incorrecto. De hecho, es posible crear una página web que viole casi todas las normas de codificación correcta HTML y el navegador seguirá mostrándola. Es evidente que esta tolerancia ha beneficiado el crecimiento de la Web, pero también ha propiciado las malas prácticas de programación hasta límites inverosímiles.

El problema es que esta tolerancia no puede trasladarse al software que funcionará en la Web semántica por dos razones: primero porque todo el edificio se apoya en XML, que a su vez, exige una codificación sistemática, rigurosa y consistente; segundo, porque los agentes que funcionarán en la Web semántica ejecutarán un tipo de procesamiento de grado mucho más fino, donde cualquier error hará inseguro su trabajo (o lo impedirá sin más).

La cantidad y la calidad de meta-información es otro de los indicadores considerados. Cuantos más y mejores metadatos contenga una página, más cerca estará el nivel de exigencia necesario para marcar semánticamente un documento y poder de este modo obtener los resultados esperados en la Web Semántica. Finalmente, la codificación en RDF es el formato para codificar metadatos adecuado para la Web Semántica y por tanto su utilización es un claro indicador de migración.

La mejora de estas tres condiciones ha de implicar cambios profundos en la actitud de los autores y editores de documentación científico-técnica. Hasta ahora el eje central en la creación de un documento digital era conseguir un buen resultado visual. Con la Web Semántica se deberá conseguir sobretodo un buen código fuente asumiéndose mayores responsabilidades en la creación de meta-información.

## **7.1 Análisis**

A continuación mostraremos un diseño de análisis experimental que se llevó a cabo a fines de 2005 utilizando los tres indicadores mencionados para analizar cuatro grupos temáticos de sedes Web y determinar en qué grado están evolucionando hacia la Web Semántica. Los grupos de sedes Web analizados para este análisis son (ver anexo para la relación completa de sitios):

- WebLogs en español sobre Biblioteconomía y Documentación
- Revistas científicas electrónicas de libre acceso de Biblioteconomía y Documentación
- Sedes web de parlamentos de países la Unión Europea

Cada grupo contenía en 2005 entre 30 y 60 sedes web y se analizaron entre 5 y 30 páginas de cada Web. Entendemos que los resultados obtenidos son difícilmente extrapolables a otros ámbitos que no sean el propio grupo analizado. Igualmente, a falta de más datos, los tres grupos analizados no tienen porqué constituir una muestra representativa para realizar afirmaciones

concluyentes sobre la Web en general. No obstante, los datos son un buen indicador sobre las dificultades reales que debe afrontar la futura migración hacia la Web semántica en aspectos *aparentemente* simples como los señalados en 5.1. donde ni siquiera entra a formar parte el mundo de las ontologías.

Los análisis se realizaron en diciembre el 2005 por medio de *DigiDocSpider*, un programa desarrollado para obtener datos de forma automática para investigaciones que evalúen la calidad de sedes web. *DigiDocSpider* ha sido creado por el grupo de investigación DigiDoc del Instituto Universitario de Lingüística Aplicada de la Universidad Pompeu Fabra en el contexto del proyecto HUM2004-03162/FILO del Plan Nacional I+D+I del Ministerio de Educación y Ciencia (España). Las tablas siguientes muestran los principales datos obtenidos en relación a los tres ejes ya señalados: calidad del código fuente, uso de metadatos y uso de RDF:

Tabla 9: Calidad del código fuente

	Revistas	WebLogs	Parlamentos	Ayunt.
Promedio por página de errores (x)html según W3C	33,02	57,78	59,8	60,94
Promedio por página de errores de accesibilidad según TAW (nivel 1)	7,82	3,25	11,26	24,03
Promedio por página de errores CSS según W3C	8,46	1,56	6,65	15,16
Total de errores (promedio por página)	49,3	62,59	77,71	100,13
Promedio por página de etiquetas desaconsejadas por W3C	47,85	16,92	74,99	73,83
Porcentaje de páginas sin DOCTYPE	55%	16%	27%	47%
Porcentaje de páginas sin xmlns en la etiqueta html	95%	27%	88%	95%
Porcentaje de páginas sin lang en la etiqueta html	92%	55%	63%	76%
Promedio por página de atributos sin comillas	32,27	15,28	35,69	52,04
Porcentaje de etiquetas en mayúsculas	19%	1%	24%	10%
Porcentaje de etiquetas breaks sin cerrar	84%	18%	82%	88%
Porcentaje de etiquetas de gráficos sin cerrar	87%	25%	90%	90%

Como se puede ver, el número de errores en el código (X)HTML que presentaban este grupo de páginas era muy alto, entre 33 y 59 errores por página de promedio. El número de errores en el código CSS era más moderado, entre 1 y 6 de promedio por página, finalmente los errores de accesibilidad de nivel 1 (TAW) fueron entre 3 y 11. En la tabla se puede observar algunos motivos de los errores del código. Por ejemplo, en el grupo de revistas digitales hay un porcentaje superior al 50% de páginas sin la declaración de DOCTYPE. Por otro lado, hay indicadores que muestran de manera muy clara que en la inmensa mayoría de ocasiones estamos ante un código que no es XHTML, sino HTML y sin depurar (etiquetas en minúsculas, etiquetas sin cerrar, atributos sin comillas, etiqueta HTML sin parámetros y gran número de etiquetas desaconsejadas).

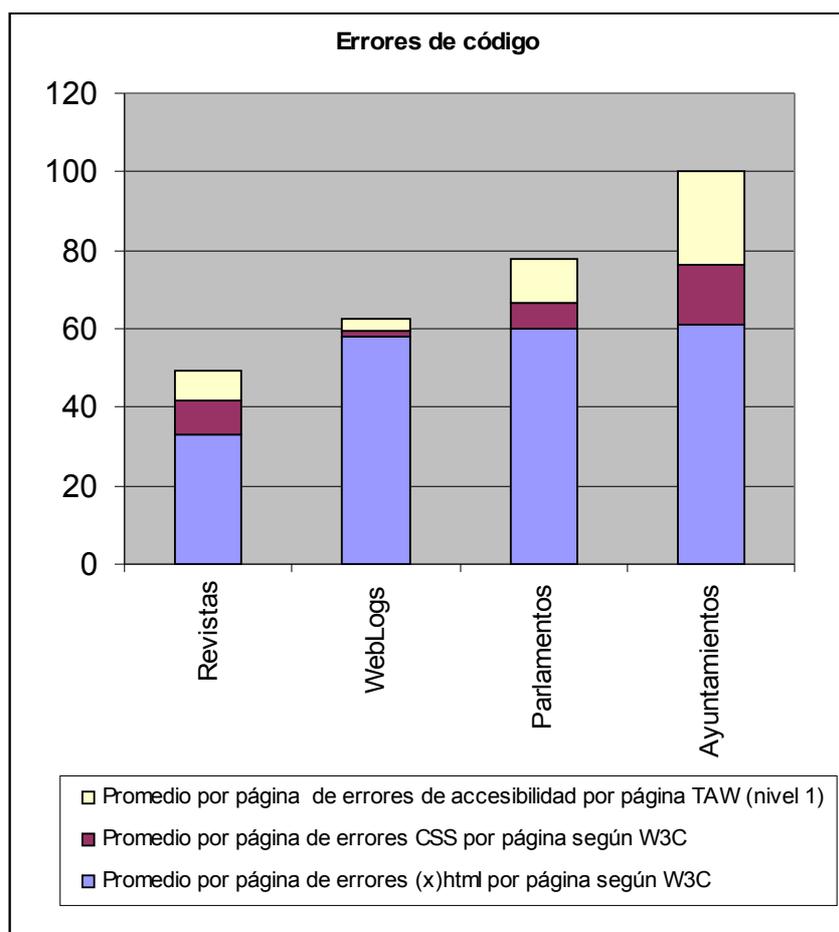


Figura 13: Promedio por página de errores de código

Tabla 10: Uso de metadatos

	WebLogs	Revistas	Parlamentos	Ayuntamientos
Promedio de etiquetas de metadatos por página	5,51	4,24	3,48	3,68
Promedio de etiquetas DC por página	1,87	1,63	0,18	0,06

El número de etiquetas de metadatos resultó ser muy bajo, estando situado entre 3 y 5 de promedio por página. Además, la mayoría correspondían a metadatos estructurales (software generador, fechas de creación o modificación, content-type...) que fueron creados automáticamente por los programas de edición de páginas. La meta-información basada en Dublin Core todavía es más escasa, entre el 0,8 y el 1,8 de promedio por página. Estos datos nos llevan a afirmar que, a excepción de casos muy concretos, en ninguno de los grupos analizados se detectó una clara política de generación de metainformación.

Tabla 11: Uso de RDF

	Revistas	WebLogs	Parlamentos	Ayuntamientos
Promedio de etiquetas RDF en la página HTML	0,016	1,042	0	0
Promedio de enlaces a ficheros RDF	0	0,003	0	0
Sedes web con etiquetas RDF en la página HTML	3,4%	32%	0%	0%
Sedes web con enlaces a ficheros RDF	0%	2%	0%	0%

Tabla 3: RDF (elaboración propia)

No se detectó la presencia de código RDF de forma significativa, ni en el interior de ficheros (X)HTML (práctica que en todo caso, el consorcio Dublín Core desaconseja) ni en un fichero vinculado con extensión .rdf. Tan solo en el grupo de WebLogs se obtuvieron datos destacables. Se observó aquí un promedio de una sola etiqueta RDF por página y un 32% de sedes web analizadas con metadatos RDF, en todo caso cifras insuficientes para dar soporte al procesamiento necesario para la Web Semántica. En parlamentos y ayuntamientos no se obtuvo ninguna codificación de meta datos en RDF. Finalmente, en el grupo de revistas tan solo 2 sedes Web (3,4%) presentaban alguna etiqueta RDF con lo que su presencia es prácticamente nula.

El análisis de estos datos se puede resumir en tres afirmaciones:

- Existía un alto nivel de errores de codificación
- Los metadatos eran realmente escasos
- No existía prácticamente codificación RDF

El análisis confirmó que en las webs analizadas todavía no se había iniciado la migración hacia la Web Semántica. Lo cierto es que difícilmente se producirá esta migración si no hay una presión por parte de servicios avanzados que necesiten datos marcados semánticamente y, al mismo tiempo, es difícil que aparezcan este tipo de servicios sin la garantía que las sedes web van a contener la meta-información semántica necesaria.

Estamos, por tanto, ante la típica situación del pez que se muerde la cola. Será necesario estar atento para detectar si en el futuro, esperemos que a corto plazo, se rompa el bucle en algún ámbito concreto, más allá de intranets o pequeños experimentos universitarios.

## 8 Conclusiones

Podemos decir que parte del programa de la Web semántica ha sido un éxito. Estamos pensando en la gran penetración de XML, cuyas consecuencias tendrán sin duda una influencia duradera en el mundo de la información en general y de la documentación en particular. Esta influencia será prácticamente invisible, pero sus consecuencias serán de enorme calado, porque significa el triunfo de aquello que siempre ha defendido la documentación: un modelo de datos no propietario, independiente de la plataforma y que garantiza la compatibilidad con futuros sistemas.

Hay otro aspecto en el cual el proyecto de la Web semántica, por así decirlo, ha sido ya un éxito: se trata de que este proyecto ha sido capaz de obligar a las empresas de software más importantes del mundo (léase Microsoft, IBM, Sun, Adobe, etc.) a tomarse en serio la importancia de la normalización.

La Web semántica puede apuntarse otros méritos: ha vuelto a poner en la agenda los estudios de Inteligencia Artificial (aunque sea “de tapadillo”), ha dado un importante soporte a la idea de

los metadatos, tal y como la entendemos los documentalistas, como demuestra el hecho de que Dublin Core haya sido adoptado como uno de los principales estándares de RDF.

De paso, ha puesto de manifiesto la necesidad de adoptar una estrategia que también es característica de las Ciencias de la Documentación: si queremos controlar un gran volumen de información es necesario sistematizar tanto el tratamiento de los datos como el de los metadatos. Nos referimos a la idea de que es necesario crear lenguajes de marcado específico que ayuden a explotar la semántica incluida en los textos, algo que la documentación ha defendido (y practicado) siempre a través de normas como ISBD, el uso de las bases de datos documentales con diccionarios de datos rigurosos y con la aplicación de lenguajes documentales a la representación del contenido de la información.

En el lado oscuro, hemos de situar la presunción de que a las alturas del 2010, si no antes, los ordenadores conectados a la Web ya serán capaces de efectuar razonamientos inteligentes gracias a la infraestructura de la Web semántica (Berners-Lee y otros, 2001). Esta es, sin duda, la peor cara de la Web semántica: su empeño en fingir que, sin necesidad de cambios profundos en las ciencias de la computación, los ordenadores inteligentes están al alcance de la mano. Su otro empeño paralelo no es mucho mejor. Consiste en fingir que, sin ningún aliciente tangible, los millones de creadores de páginas web de todo el planeta, de una forma espontánea y sin embargo coordinada, empezarán a crear páginas web pobladas con metadatos que serán a la vez fiables, creíbles, honestos y ¡técnicamente impecables! y todo ello con el uso de estándares y tecnologías cuya comprensión dista de ser trivial, a saber: XML, Dublin Core, RDF, RDF Schema, OWL, etc.

Al inicio de este capítulo hemos presentado la Web semántica según dos visiones: la de la Inteligencia Artificial y la del Procesamiento robusto. Creemos que la visión de la Inteligencia Artificial es la parte débil de la Web semántica y la que puede poner en peligro su credibilidad a medio plazo. Por suerte, también creemos que la visión del Procesamiento robusto justifica por sí sola todo el proyecto de la Web semántica. Dicho de otra manera: si tuviéramos razón (y preferiríamos equivocarnos en este punto) y la visión de la Inteligencia Artificial estuviera condenada al fracaso, bastarían los enormes beneficios de las aportaciones de la visión del Procesamiento robusto para dar todo nuestro apoyo a este proyecto.

Una de las grandes contradicciones de la Web semántica es que, aunque a nivel social proporciona proporcionar grandes beneficios, a nivel individual no aporta ninguno. ¿A qué nos referimos? Ni más ni menos que al siguiente hecho: sin duda, a nivel social sería muy beneficioso que todas y cada una de las páginas publicadas en la Web tuvieran un código fuente impecable; pero a nivel individual no hay ninguna recompensa por preocuparse por ello. De facto, debido al negligente funcionamiento de la mayor parte de editores de páginas web de tipo WISIWYG no solo es mucho más fácil hacer una página no estándar, sino que resulta directamente imposible crear código fuente estándar. Es más, aunque autores y webmasters se esfuercen en ello, no existe ninguna compensación tangible para ese esfuerzo adicional.

Por tanto, y aunque sea con mentalidad de ONG, ¿qué podemos hacer a favor de la Web semántica si creemos en sus beneficios a escala social aunque, por ahora, sepamos que aporta escasos o ningún beneficio individual? Los organismos vinculados al mundo de la documentación y la promoción del conocimiento (archivos, bibliotecas, centros de documentación, CRAI, etc.) y los organismos vinculados a la su creación y difusión (universidades, centros de investigación, museos, institutos de investigación, etc.) deberían sentirse obligados por la visión de la Web semántica debido a la naturaleza intrínseca de su misión, a saber: promocionar y difundir el conocimiento y favorecer la creación de nuevo conocimiento desde el punto de vista de su beneficio social.

Por tanto, al menos a corto plazo, esta clase de organizaciones debería sentirse obligados a: (1) el uso exclusivo de estándares en sus sitios web, lo que significa utilizar exclusivamente

XHTML y CSS en sus sitios y desterrar cualquier tipo de codificación no estándar; (2) comprometerse en una política constante de mejora de la calidad del código fuente de sus sitios web; (3), diseñar y aplicar políticas de metadatos basadas en RDF/XML, Dublin Core y/o cualesquiera otro sistema de metadatos compatible con el proyecto de la Web semántica (Skos-Core, MODS, EAD, etc.).

Por último, podemos preguntarnos, *¿qué enseña la Web semántica a los documentalistas?* En nuestra opinión, nos enseña algo que se supone que ya sabíamos, esto es: si tomamos un conjunto de datos y los etiquetamos sistemática y cuidadosamente, tenemos lo más parecido a la inteligencia (sin ser inteligencia). Si las bases de datos exhiben un notable grado de inteligencia en comparación con la Web es porque en una base de datos, todos los datos están "etiquetados" o sea, forman parte de los valores admitidos de un campo predefinido. Cada campo, a su vez, tiene unas propiedades o restricciones bien definidas: se trata de un campo de texto, o de un campo numérico, o lógico, etc. Por último, todos los datos en una base de datos están sistematizados: cada registro responde a la misma estructura, así que la mera posición (la sintaxis) genera sentido (semántica). Así que, lo que es (genialmente) nuevo en la Web semántica es la idea de convertir toda la Web es la más gigantesca base de datos documental que la humanidad pudiera haber soñado jamás. Demos todo nuestro apoyo a este proyecto.

## 9 Bibliografía

Abadal, E.; Codina L. 2005. *Bases de datos documentales: características, funciones y método*. Madrid: Síntesis, 220 p.

Alesso, H. P.; Smith, C. F. 2005. *Developing semantic web services*. Wlley: AK Peters, 445 p.

Antoniou, G.; Van Harmelen, F. 2004. *A semantic web primer*. Cambridge: MIT, 238 p.

Berners-Lee, T. *Semantic Web Architecture*. Accesible en:  
<http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide11-0.html>

Berners-Lee, T.; Hendler, J.; Lassila, O. 2001. "The Semantic Web". *Scientific American*, May 2001 (Accesible en formato html en la web de *Scientific American*:  
[www.scientificamerican.com](http://www.scientificamerican.com) )

Brachman, R. J.; Levesque, H. J. (2004). *Knowledge representation and reasoning*. San Francisco: Morgan Kaufmann, 381 p.

Chowdhury, G. Gobinda 2004. *Introduction to modern information retrieval*. (Second edition) London: Facet, 437 p.

Copeland, J. 1996. *Inteligencia artificial: una introducción filosófica*. Madrid: Alianza, 1996, 421 p.

Daconta, M.C.; Obrst, L.J.; Smith, K.T. 2003. *The semantic web*. Indianapolis: Wiley, 281 p.

Davies, J.; Fensel, D.; Van Harmelen, F. 2003. *Towards the semantic web: Ontology-drives knowledge management*. Chichester: John Wiley, 288 p.

Grubber, T. 1993. "A translation approach to portable ontologies". *Knowledge Acquisition*, Vol. 5, Issue 2, pp. 199-220

Harold, E.R.; Means, S. 2005. *XML imprescindible*. Madrid: Anaya, 832 p.

- Miller, D.R.; Clarke, K.S. 2004. *Putting XML to work in the library: Tools for improving access and management*. Chicago: American Library Association, 205 p.
- Lacy, Lee W. 2004. *OWL: Representing information using the Web Ontology Language*. Victoria, BC: Trafford, 282 p.
- Mercer, Dave. 2002. *Fundamentos de programación en XML*. Bogotá: McGraw-Hill, 436 p.
- Miles, A.; Brickley, D. 2005. "SKOS Core Guide". *World Wide Web Consortium*, November 2005. Acceso: <http://www.w3.org/TR/2005/WD-swbp-skos-core-guide-20051102/>
- Passin, T. B. 2004. *Explorer's guide to the semantic web*. Greenwich: Manning, 281 p.
- Penrose, R. 1991. *La nueva mente del emperador*. Madrid: Mondadori, 597 p.
- Powers, S. 2003. *Practical RDF*. Sebastopol: O'Reilly, 331 p.
- Rodríguez, Alonso. 2004. *Publicación en Internet y tecnología XML*. Madrid: RA-MA, 2004, 438 p.
- Senso, J. A. 2004. "Resource Description Framework". En: Rovira, C.; Codina, L. (dir.). *Anuario Hipertext.net*, n. 1, 2004. Acceso: <http://www.hipertext.net/web/pag236.htm>
- Senso, J. A. 2003. "Herramientas para trabajar con rdf". *El profesional de la información*, vol. 12, n. 2, 2003, pags. 132-139
- Tramullas, J. 2001. "La recuperación de información en el World Wide Web". En J. López Yepes (coord.) *Manual de Ciencias de la Documentación*. Madrid: Síntesis, 2001, p. 601-631
- World Wide Web Consortium. 2006. "Semantic Web". *Semantic Web Activity*, Acceso: <http://www.w3.org/2001/sw/>

# 10 Anexo: grupos de sedes web analizados

## 1) Grupo 1: WebLogs

**Criterio de selección:** WebLogs en español sobre Biblioteconomía y Documentación

**Fecha del análisis:** entre el 7/12/2005 y el 10/12/2005

**Número de sedes web analizadas:** 60

**Número de páginas analizadas en cada sede web:** 11

**Sedes web analizadas:** ¡Incatalogable (<http://excatedra.blogspot.com/>), A nosa biblioteca IES As Mariñas (<http://biblioweb.blogspot.com/>), A pie de página (<http://apiepagina.bitacor.com/>), A propósito de (<http://www.apropositode.blogspot.com/>), Alas y balas (<http://www.alasybalas.com/>), ALFIN: Avances de la alfabetización Informacional (<http://alfin.blogspot.com/>), Amoxcalli Blog (<http://amoxcalli.blogspot.com/>), Baiget Info Notas (<http://baiget.blogspot.com/>), Bauen (<http://bauen.lamevaweb.info/>), Biblioblog. Bitácora sobre bibliotecas (<http://www.interreg-eet.info/weblogs/biblioblog/index.html>), Bibliofreesoft - resources (<http://clopezpe.blogspot.com/>), Bibliometría... infometría, ciencias, cibermetría (<http://www.bibliometria.com/>), Biblio's (<http://libiblios.blogspot.com/>), Biblioteca telecomunicación Madrid (<http://rarebookcollection.bitacor.com/>), Bibliotecosas (<http://bibliotecosas.blogia.com/>), Bienvenido te colocó (<http://bienvenido.blogspot.com/>), blog de Libros y bits (<http://jamillan.com/lbblog6.htm>), Bradomin 4.0 Mi brújula personal en la Red (<http://bradomin.iespana.es/>), Busca noticias (<http://buscanoticias.blogspot.com/>), Buscar para encontrar buscadores, audiovisual y todo lo demás (<http://buscarencontrar.blogspot.com/>), Catorze.blog (<http://www.catorze.com/blog/>), Chacobo (<http://www.chacobo.com/>), Deakialli (<http://deakialli.bitacor.com/>), Dejavu (<http://www-dejavu.blogspot.com/>), Diario de la realización de un libro medieval (<http://migranlibro.bitacor.com/>), Documenta2 (<http://documenta2.blogspot.com/>), Documentación (<http://www.documentacion.com.es/>), Documentación informativa (<http://fvalle.blogspot.com/>), Documentación sin límites (<http://docunlimited.blogspot.com/>), Documentación, biblioteconomía e información (<http://www.lacocelera.com/documentacion>), Documental (<http://documental.blogspot.com/>), Documental Mente (<http://documentalmente.blogspot.com/>), Dokulan (<http://dokulan.blogspot.com/>), dokumenta (<http://dokumenta.blogia.com/>), El documentalista audaz (<http://sirmijail.blogspot.com/>), El hombre irrazonable (<http://victormacias.blogia.com/>), Evaluación de la Ciencia y la Comunicación Científica (<http://ec3noticias.blogspot.com/>), Exlibris libros y bibliotecas conocimientos y recuerdos (<http://gorqui.blogspot.com/>), Expediente, Web, Blog y otros documentos del montón (<http://documentalqueridowatson.blogspot.com/>), Fall in blog (<http://fallinblog.webcindario.com/>), GeekTeca (<http://geekteca.blogspot.com/>), Guía de buscadores blog (<http://www.guiabuscadores.com/blog/>), Infoesfera (<http://www.infoesfera.com/>), Infogestores (<http://www.infogestores.cl/>), Joana Albret Biblioteconomía Mintegia (<http://www.eibar.org/blogak/joana-albret/>), K~Utopía: El nuevo mundo de la Documentación (<http://anikautopia.blogia.com/>), La librería pequeños retales de literatura (<http://www.lalibreria.blogspot.com/>), Métodos de Busca (<http://www.metodosdebusca.com/118/index.php>), Mòbils BiD (<http://mobilsbid.blogspot.com/>), Outlook. Diseño web, usabilidad y Documentación (<http://outbook.blogspot.com/>), Programación Lógica y Recuperación de Información (<http://programacionlogica.blogspot.com/>), R020 Biblioteconomía y ciencias de la información (<http://www.r020.com.ar/>), Recuperación de Información en la Web (<http://irsweb.blogspot.com/>), Red de bibliotecas CEP (<http://bibliored.bitacor.com/>), Servicio de alerta (<http://docdigital.typepad.com/>), Sistema de Archivos (<http://sistemadearchivos.blogspot.com/>), TekaTeka (<http://tekateka.blogspot.com/>), Telendro (<http://telendro.com.es/>), Véase además (<http://www.veaseademas.com/>), Visto y Leído (<http://vistoyleido.blogspot.com/>).

## 2) Grupo 2: Revistas electrónicas

**Criterios de selección:** revistas científicas electrónicas de libre acceso de Biblioteconomía y Documentación

**Fecha del análisis:** entre el 13/11/2005 y el 10/12/2005

**Número de sedes web analizadas:** 60

**Número de páginas analizadas en cada sede web:** 31

**Sedes web analizadas:** ACIMED (<http://scielo.sld.cu/scielo.php>), Anales de Documentación (<http://www.um.es/fccd/anales/>), Ariadne (<http://www.ariadne.ac.uk/>), Arquivística.net (<http://www.arquivistica.net/ojs/index.php>), Australian Academic and Research Libraries (<http://alia.org.au/publishing/aarl/index.html>), Avisos. Noticias de la Real Biblioteca (<http://www.patrimonionacional.es/realbiblioteca/>), Biblios: Revista Electrónica de Biblioteconomía y Ciencias de la Información (<http://www.bibliosperu.com/>), Biblioteca Universitaria (<http://dgb.unam.mx/revista/Revista%20online/>), Bibliotime Rivista elettronica per le biblioteche (<http://didattica.spbo.unibo.it/bibliotime/>), BiD: textos universitaris de Biblioteconomía i Documentació (<http://www.ub.es/bid/>), Biomedical Digital Libraries (<http://www.bio-diglib.com/home/>), Boletín de la Asociación Andaluza de Bibliotecarios (<http://www.aab.es/puboletin.html>), Boletín RedIris (<http://www.rediris.es/rediris/boletin/>), Bulletin des Bibliothèques de France (<http://bbf.ensib.fr/>), Bulletin of the American Society for Information Science and Technology (<http://www.asis.org/Bulletin/index.html>), Ciência da Informação (<http://www.scielo.br/scielo.php>), Clip boletín de la SEDIC ([http://www.sedic.es/p\\_boletinclip.htm](http://www.sedic.es/p_boletinclip.htm)), Correo Bibliotecario (<http://www.bcl.jcyl.es/correo/index.php>), Cuadernos de documentación multimedia (<http://multidoc.rediris.es/cdm/index.php>), Cybermetrics (<http://www.cindoc.csic.es/cybermetrics/vol9iss1.html>), DIGITHUM (<http://www.uoc.edu/digithum/>), D-Lib Magazine (<http://www.dlib.org/>), Electronic Journal of

Information Systems in Developing Countries (<http://new.ejisdc.org/ojs/>), Electronic Journal of Knowledge Management (<http://www.ejkm.com/>), Encontros Bibli (<http://www.encontros-bibli.ufsc.br/sumario.htm>), Exploit Interactive (<http://www.exploit-lib.org/>), First Monday (<http://www.firstmonday.org/>), High Energy Physics Libraries Webzine (<http://library.cern.ch/HEPLW/>), hipertext.net (<http://www.hipertext.net/web/pag100.htm>), Human IT (<http://www.hb.se/bhs/ith/humanit.htm>), IFLA Journal (<http://www.ifla.org/V/iflaj/>), Information - Interaction - Intelligence (<http://www.revue-i3.org/index.html>), Information Research: an international electronic journal (<http://informationr.net/ir/>), Information Technology and Disabilities (<http://www.rit.edu/~easi/itd.htm>), Informing Science (<http://inform.nu/>), International Review of Information Ethics (<http://www.i-r-i-e.net/>), Investigación bibliotecológica ([http://www.ejournal.unam.mx/iibiblio/iib\\_v18-38.html](http://www.ejournal.unam.mx/iibiblio/iib_v18-38.html)), Issues in Science and Technology Librarianship (<http://www.istl.org/>), IT&Society (<http://www.stanford.edu/group/siqss/itandsociety/>), Journal of Digital Information (<http://jodi.tamu.edu/>), Journal of Electronic Publishing (<http://www.press.umich.edu/jep/>), Journal of eLiteracy (<http://www.jelit.org/>), Journal of Information, Law & Technology (<http://www2.warwick.ac.uk/fac/soc/law/elj/jilt/index.html>), Journal of Issues in Informing Science and Information Technology (<http://iisit.org/>), Journal of Librarianship and Information Science (<http://lis.sagepub.com/>), Journal of the Medical Library Association (<http://www.pubmedcentral.gov/tocrender.fcgi?journal=93>), Liber. Revista de bibliotecología ([http://www.ambac.org.mx/publicaciones/liber1\\_4\\_2004.htm](http://www.ambac.org.mx/publicaciones/liber1_4_2004.htm)), Library Philosophy and Practice (<http://www.webpages.uidaho.edu/~mbolin/lpp.htm>), Libres: Library and Information Science Research Electronic Journal (<http://libres.curtin.edu.au/>), Pez de Plata: Bibliotecas Públicas a la Vanguardia (<http://www.pezdeplata.org/>), Revista ACB (<http://www.acbsc.org.br/revista/ojs/index.php>), Revue électronique suisse de science de l'information (<http://campus.hesge.ch/ressi/welcome.html>), School Library Media Research (<http://www.ala.org/ala/aasl/aaslpubsandjournals/slmlr/schoollibrary.htm>), Simbiosis: Revista electrónica de ciencias de la información (<http://simbiosis.uprrp.edu/index.htm>), SIMILE. Studies in Media & Information Literacy Education (<http://www.utpjournals.com/jour.ihml?lp=simile/simile.html>), Sociedad de la Información (<http://www.sociedadelainformacion.com/>), Solaris Information Communication (<http://biblio-fr.info.unicaen.fr/bnum/jelec/Solaris/>), The Bonefolder, an e-journal for the bookbinder and book artist (<http://www.philobiblon.com/bonefolder/index.html>), The Electronic Journal of Academic and Special Librarianship (<http://southernlibrarianship.icaap.org/>), The Library Quarterly (<http://www.journals.uchicago.edu/LQ/journal/home.html>).

### 3) Grupo 2: Parlamentos

**Criterios de selección:** sedes web de parlamentos de países la Unión Europea

**Fecha del análisis:** entre el 11/12/2005 y el 12/12/2005

**Número de sedes web analizadas:** 33

**Número de páginas analizadas en cada sede web:** 11

**Sedes web analizadas:** Alemania (Bundesrat (<http://www.bundesrat.de/Site/Inhalt/DE/>), Alemania (Bundestag (<http://www.bundestag.de/>), Austria (<http://www.parlament.gv.at/>), Bélgica (Cámara de representantes ([http://www.dekamer.be/kvvr/choose\\_language.cfm](http://www.dekamer.be/kvvr/choose_language.cfm)), Bélgica (Senado) (<http://www.senate.be/>), Chipre (<http://www.parliament.cy/>), Dinamarca (<http://www.folketinget.dk/>), Eslovaquia (<http://www.nrsr.sk/>), Eslovenia (<http://www.dz-rs.si/>), España (congreso de Diputados) (<http://www.congreso.es/>), España (Senado) (<http://www.senado.es/>), Estonia (<http://www.riigikogu.ee/>), Finlandia (<http://www.eduskunta.fi/index.html>), Francia (Asamblea Nacional) (<http://www.assemblee-nationale.fr/>), Francia (senado) (<http://www.senat.fr/>), Grecia (<http://www.parliament.gr/english/default.asp>), Holanda (Cámara de Representantes) (<http://www.tweedekamer.nl/>), Holanda (Senado) (<http://www.eerstekamer.nl/>), Hungría (<http://www.parlament.hu/>), Irlanda (<http://www.oireachtas.ie/>), Italia (Cámara de Diputados) (<http://www.camera.it/>), Italia (Senado) (<http://www.senato.it/>), Letonia (<http://www.saeima.lv/>), Lituania (<http://www.lrs.lt/>), Luxemburgo (<http://www.chd.lu/default.jsp>), Polonia (Sejm) (<http://www.sejm.gov.pl/>), Polonia (Senado) (<http://www.senat.gov.pl/>), Portugal (<http://www.parlamento.pt/>), Reino Unido (Cámara de los Loes) (<http://www.parliament.uk>), Reino Unido (Cámara de los Comunes) (<http://www.parliament.uk/>), República Checa (Cámara de Diputados) (<http://www.psp.cz/>), República Checa (Senado) (<http://www.senat.cz/>), Suecia (<http://www.riksdagen.se/>).

### 4) Grupo 2: Ayuntamientos

**Criterios de selección:** sedes web de ayuntamientos de Cataluña de más de 20.000 habitantes

**Fecha del análisis:** el 18/12/2005

**Número de sedes web analizadas:** 55

**Número de páginas analizadas en cada sede web:** 5

**Sedes web analizadas:** Ajuntament de Badalona (<http://www.aj-badalona.es>), Ajuntament de Barberà del Vallès (<http://www.bdv.es>), Ajuntament de Barcelona (<http://www.bcn.es/>), Ajuntament de Blanes (<http://www.blanes.net>), Ajuntament de Cambrils (<http://www.cambrils.altanet.org>), Ajuntament de Castelldefels (<http://www.castelldefels.org/>), Ajuntament de Cerdanyola del Vallès (<http://www.cerdanyola.org>), Ajuntament de Cornellà de Llobregat (<http://www.cornellaweb.com>), Ajuntament de Figueres (<http://www.figueresciutat.com>), Ajuntament de Gavà (<http://www.gavaciutat.net>), Ajuntament de Girona (<http://www.ajuntament.gi/>), Ajuntament de Granollers (<http://www.granollers.org/>), Ajuntament de l'Hospitalet de Llobregat (<http://www.l-h.net>), Ajuntament de Lleida (<http://www.paeria.es/>), Ajuntament de Lloret de Mar (<http://www.lloret.org>), Ajuntament de Manresa (<http://www.ajmanresa.org>), Ajuntament de Martorell (<http://www.ajmartorell.org>), Ajuntament de Mataró (<http://www.mataro.org>), Ajuntament de Molins de Rei (<http://www.molinsderei.net/ajuntament>), Ajuntament de

Mollet del Vallès (<http://www.molletvalles.net/>), Ajuntament de Montcada i Reixac (<http://www.montcada.org>), Ajuntament de Pineda de Mar (<http://www.pinedademar.org>), Ajuntament de Premià de Mar (<http://www.premiademar.org>), Ajuntament de Reus (<http://www.reus.net>), Ajuntament de Ripollet (<http://www.ripollet.org>), Ajuntament de Rubí (<http://www.rubiciutat.net>), Ajuntament de Sabadell (<http://www.ajsabadell.net/>), Ajuntament de Salt (<http://www.viladesalt.org/>), Ajuntament de Sant Adrià de Besòs (<http://www.sant-adria.net>), Ajuntament de Sant Andreu de la Barca (<http://www.stabarca.com/>), Ajuntament de Sant Boi de Llobregat (<http://www.stboi.es/>), Ajuntament de Sant Cugat del Vallès (<http://www.santcugatobert.net/>), Ajuntament de Sant Feliu de Llobregat (<http://www.santfeliu.org/>), Ajuntament de Sant Joan Despí (<http://www.sjdespi.com/>), Ajuntament de Sant Pere de Ribes (<http://www.santperederibes.org>), Ajuntament de Sant Vicenç dels Horts (<http://www.svh.es/>), Ajuntament de Santa Coloma de Gramenet (<http://www.grame.net>), Ajuntament de Santa Perpètua de Mogoda (<http://www.santaperpetua.diba.es>), Ajuntament de Sitges (<http://www.sitgestur.com/>), Ajuntament de Tarragona (<http://www.ajtarragona.es>), Ajuntament de Terrassa (<http://www.terrassa.org/>), Ajuntament de Tortosa (<http://www.tortosa.altanet.org>), Ajuntament de Valls (<http://www.ajvalls.org/>), Ajuntament de Vic (<http://www.ajvic.es/>), Ajuntament de Viladecans (<http://www.ajviladecans.es/>), Ajuntament de Vilafranca del Penedès (<http://www.ajvilafranca.es/>), Ajuntament de Vilanova i la Geltrú (<http://www.vilanova.org/>), Ajuntament del Masnou (<http://www.elmasnou.net>), Ajuntament del Prat de Llobregat (<http://www.aj-elprat.es/>), Ajuntament del Vendrell (<http://www.elvendrell.net>), Ajuntament d'Esparreguera (<http://esparreguera.org>), Ajuntament d'Esplugues de Llobregat (<http://www.ajesplugues.es/>), Ajuntament d'Igualada (<http://www.aj-igualada.net>), Ajuntament d'Olesa de Montserrat (<http://www.olesam.org>), Ajuntament d'Olot (<http://www.olot.org>).

---