

Web semántica: un nuevo enfoque para la organización y recuperación de información en el web

[Lic. Keilyn Rodríguez Perojo¹](#) y [Lic. Rodrigo Ronda León²](#)

RESUMEN

Se realiza un acercamiento teórico-conceptual al desarrollo de la organización y recuperación de la información en ambiente automático, mediante la indización y la clasificación automática de la información como procesos previos a la recuperación de información. Se analiza la evolución experimentada por los sistemas de organización de la información y el conocimiento: taxonomías, tesauros y Topic Maps hacia estructuras más complejas como las ontologías. Se examinan algunos esquemas de meta-datos, concebidos para la descripción de la información en el web, así como los componentes y tecnologías que integran el modelo teórico de la web semántica. Finalmente, se explica el modelo constructivo de capas que garantizará el funcionamiento de toda su infraestructura.

Palabras clave: Web semántica, organización y recuperación de la información, tecnologías de la información y la comunicación.

ABSTRACT

A theoretical and conceptual approach to the development of the information organization and retrieval is made in an automatic environment by means of the automatic indexing and classification of the information as processes prior to information retrieval. The evolution experienced by the systems of organization of information and knowledge is analysed by taxonomies, thesaurus and Topic Maps toward more complex structures, such as ontologies. Some metadata schemes created for the description of the information in the web, as well as the components and technologies that integrate the theoretical model of the web semantics, are examined. Finally, the constructive stratum model that will guarantee the functioning of its entire infrastructure is explained

Key words: Semantic Web, information organization and retrieval, information and communication technologies.

Copyright: © ECIMED. Contribución de acceso abierto, distribuida bajo los términos de la Licencia Creative Commons Reconocimiento-No Comercial-Compartir Igual 2.0, que permite consultar, reproducir, distribuir, comunicar públicamente y utilizar los resultados del trabajo en la práctica, así como todos sus derivados, sin propósitos comerciales y con licencia idéntica, siempre que se cite adecuadamente el autor o los autores y su fuente original.

Cita (Vancouver): Rodríguez Perojo K. Web semántica: un nuevo enfoque para la organización y la recuperación de información en el Web. Acimed 2005;13(6). Disponible en: http://bvs.sld.cu/revistas/aci/vol13_6_05/aci03605.htm Consultado: día/mes/año.

“Científicos y especialistas poco inclinados a la ciencia ficción o a la especulación futuroológica – más cerca de Nobeert Wiener, creador de la Cibernética , que de Arthur Clarke, padre de la comunicación por satélite,- anunciaron una época radicalmente distinta, basada en la microelectrónica y la transmisión global instantánea. Incluso previeron el surgimiento de nuevas formas de civilización y culturas generales por el procesamiento masivo de datos e imágenes accesibles a escala universal mediante satélites y redes digitales”

Enrique González Manet ” La Era de las Nuevas Tecnologías”

Actualmente, los sitios web emplean el HTML (*Hypertext Markup Language*) como lenguaje estándar para la representación de la información. “La evolución experimentada por el HTML desde sus inicios no sólo permite mostrar información textual sino que puede incluir también imágenes, y presentarla en un formato particular”. 1

La web semántica es un área prolífera, situada en la confluencia de la inteligencia artificial y las tecnologías web, que propone nuevas técnicas y paradigmas para la representación de la información y el conocimiento; para facilitar, tanto localizar como el compartir, integrar y recuperar recursos. 2

Dicho enfoque propone enriquecer la estructura de la información y agregar componentes semánticos que puedan procesarse de forma automática. La nueva generación de formatos está encabezada por XML (*Extensible Markup Language*) y RDF (*Resource Description Framework*), los cuales incluirán *ontologías* -taxonomía de conceptos con atributos y relaciones que proporcionan un vocabulario consensuado para definir redes semánticas de unidades de información interrelacionadas- que especificarán las reglas lógicas para que los agentes de software reconozcan y clasifiquen cada concepto. La evolución del web, en opinión de Pablo Castells, durante los últimos 15 años, no puede pasar por alto los siguientes acontecimientos: 2

1989: *Tim Berners Lee* presenta su proyecto WWW en el CERN (*Conseil Européen pour la Recherche Nucléaire*).

1993: Creación de los primeros servidores Web y el navegador *Mosaic* .

1994: Creación del Consorcio Web (World Wide Web Consortium o W3C).

1997: Creación de SHOE (*Simple HTML Ontology Extensions*), primer antecedente de la web semántica, basado en HTML.

Son numerosos los proyectos desarrollados en Internet con lenguajes de codificación de ontologías. El servidor Ontolingua, resultado del KSE (*Knowledge Sharing Effort*), ofrece herramientas para crear ontologías, integrarlas con otras existentes e incorporarlas a nuevos productos de software. Otro enfoque es el aportado por *Luke, Spector y Rager* con el desarrollo de SHOE (*Simple HTML Ontology Extensions*) como complemento semántico de HTML, que refleja el contenido de la página web y que puede utilizarse por agentes de software para el descubrimiento de información. Posteriormente SHOE ha evolucionado hacia RDF, y es OWL (*Web Ontology Language*), la más reciente especificación, mantenida por el Consorcio del Web y que intenta proveer de un lenguaje que pueda utilizarse para describir clases y relaciones entre ellas inherentes a documentos y aplicaciones web.

La utilización de ontologías también está presente en el proyecto FERMI (*Formalization and Experimentation on the Retrieval Multimedia Information*), bajo la supervisión de *C. J “Keith” van Rijsbergen*, en el que se incluyen herramientas de planificación, descubrimiento y selección de recursos de información multimedia. El proyecto IMP (*Information Manifold Project*) desarrollado en el ámbito de Bell Labs, en el que se hace uso de las ontologías para identificar las fuentes de información pertinentes a una búsqueda, acceder a ellas, obtener documentos relevantes, compararlos, seleccionar los más adecuados y ofrecer un resumen previo al usuario. En el ámbito de la medicina, se destaca el proyecto UMLS (*Unified Medical Language System*), desarrollado por la National Library of Medicine de los Estados Unidos, que utiliza las ontologías como una herramienta más para el acceso, integración y recuperación de información biomédica.

En este sentido, el grupo de trabajo SWAD Europe (*Semantic Web Advanced Development*), “tiene como objetivo poner de manifiesto, mediante ejemplos prácticos, cómo este conjunto de tecnologías suponen una ventaja real para la actual web, y resuelven problemas en áreas como: tesauros, clasificaciones, agendas, búsqueda de recursos etcétera”. 3

A partir de la integración de toda una infraestructura tecnológica, “que permita el intercambio global de conocimiento asistido por máquina”, 4 y la codificación del significado de la información mediante lenguajes de marcado, toma forma el concepto de la web semántica como “una extensión del web actual en el que el significado de la información esté bien definido, y permita al hombre y las máquinas trabajar en estrecha cooperación”. 5

La web semántica es una extensión del web cuya idea básica es tener los datos definidos y relacionados para que su uso sea más efectivo y sea posible su automatización, integración y re-utilización por medio de diferentes aplicaciones, es decir, pretende proporcionar una infraestructura que permita que las páginas web, las bases de datos, los programas y aplicaciones, los dispositivos, tanto personales como los empleados en el hogar, puedan consumir y producir datos, sin los problemas causados por los diferentes protocolos de acceso a la información que hacen de la transferencia de contenidos una tarea ardua y difícil. El soporte principal para la organización, almacenamiento y distribución de la información siguen siendo los sistemas taxonómicos, es decir, sistemas de clasificación que respondan a necesidades concretas de las empresas, bibliotecas y centros de información en favor no sólo de los usuarios humanos, sino de las máquinas también.

REPRESENTACIÓN GRÁFICA DEL CONOCIMIENTO

Taxonomías: concepto y aplicación

Cuando se busca información en las páginas amarillas de un directorio telefónico, específicamente en el índice de categorías, encontramos en los encabezamientos de Automóviles otras subcategorías:

- Automóviles, agencias de.
- Automóviles, alquiler de.
- Automóviles, reparaciones.

Ello representa un ejemplo de taxonomía; el conjunto de páginas es una gran taxonomía ordenada alfabéticamente con la finalidad adicional de asistir al usuario en la búsqueda de información. “El primer acercamiento hacia la web semántica y el uso de servicios web (*Web Services*) es la expresión de taxonomías legibles por máquina“. 6 Las taxonomías constituyen formas de clasificar y categorizar un grupo de elementos en forma de jerarquías; es simplemente una estructura en forma de árbol con ramificaciones y cada punto de estas constituye un nodo. De forma general, para la Biología una taxonomía propone “el estudio de los principios generales de la clasificación científica: en particular, la clasificación sistemática, es la clasificación ordenada de plantas y animales acorde con sus relaciones naturales”. 7 La evolución acelerada de las tecnologías de información ha provocado que el concepto de taxonomía -que hasta hace pocos años se manejaba sólo en el campo biológico- atienda no solamente a construcciones abstractas del lenguaje natural, sino también a la clasificación de entidades de información en forma de jerarquías, según las relaciones que en el mundo real ellas representan.

Así cada nodo de la taxonomía constituye una entidad de información que tiene lugar en mundo real y cada enlace entre nodos representa una relación entre clases, donde estas relaciones, a su vez, representan clases de objetos, cuya terminología se le atribuye a la programación orientada a objetos en

Informática.

En la medida que se asciende o desciende en determinada jerarquía, las taxonomías adquieren un grado mayor o menor de generalización o especificación. En la clasificación de información, permiten establecer relaciones simples o compuestas para un espacio de información -esquemas de meta-datos, tesauros, modelos conceptuales, Topic Maps y ontologías.

Desde este punto de vista, una taxonomía es una jerarquía semántica en la que las entidades de información se relacionan mediante clases y subclases; la primera es semánticamente más fuerte que la segunda, y por ello se enfatiza en las taxonomías semánticamente débiles y semánticamente fuertes. Las taxonomías del primer orden carecen de complejidad para expresar agudeza o riqueza en el significado, pero las del segundo orden tratan de utilizar la noción de propiedades o atributos para diferenciar una subclase de la clase superior.

El uso más frecuente de las taxonomías en ambiente web es en la navegación, especialmente cuando se tiene una idea general de lo que se busca mediante la consulta a motores de búsquedas o directorios web. La Clasificación Decimal de Dewey, en su versión electrónica, es una taxonomía utilizada en las bibliotecas digitales para proporcionarle a los lectores, temáticas o tópicos generales sobre los cuales tratan sus materiales de consulta. El Sistema de Clasificación Industrial Internacional Estándar (NAICS) , concebida como taxonomía para el comercio electrónico y desarrollada por los Estados Unidos, Canadá y México, posibilita comparar estadísticas sobre la actividad empresarial en todo Norte América.

Las taxonomías, como los tesauros y las ontologías, son mecanismos que permiten estructurar datos e información para dotar a los sistemas de recuperación de información de al menos, un mínimo de semántica. En el web, pueden utilizarse para ayudar a los usuarios/clientes a encontrar productos y servicios, así como para representar mejor sus necesidades de información. Como estructuras enfocadas a la clasificación de contenido, posibilitan que los motores de búsquedas y otras aplicaciones que utilizan taxonomías como los servicios web (*Web Services*), localicen entidades de información más rápido y con mayor grado de precisión y exhaustividad mediante UDDI (*Universal Description Discovery and Integration*).

Las taxonomías son excelentes mecanismos para clasificar entidades de información, de ahí que las ontologías utilizan taxonomías como columna vertebral para su funcionamiento como base estructural. A juicio de *Taylor* , las ontologías, que en el campo de la recuperación de información y la inteligencia artificial suponen un avance en la interrelación entre los humanos y las computadoras, pueden ser no lingüísticas -empleadas para la creación de agentes inteligentes- y lingüísticas, al vincularse con aspectos gramáticos, semánticos y sintácticos.

En unos casos, estas ontologías lingüísticas se reducen a una lista jerárquica de términos en un área específica o dominio del conocimiento y en otros, son vocabularios controlados con categorías que incluyen un análisis semántico de palabras para su posterior categorización y enlace con otras, en

términos similares a las relaciones propias de los tesauros y mapas conceptuales.

Mapas conceptuales

Los mapas conceptuales surgen en ámbito de la didáctica de las disciplinas científicas de la mano de *Joseph Novak* en 1984. Dicho autor propone el uso de una herramienta que llama *Concept Map* y que define como “un dispositivo esquemático que representa un conjunto de significados conceptuales incluidos en una estructura de proposiciones”.⁸ Para *Novak*, los mapas conceptuales no deben considerarse herramientas para el aprendizaje, sino un medio para desarrollar las actividades y rutinas necesarias para que este se produzca.

Los mapas conceptuales no aplican ningún control del vocabulario. Para su construcción no existe ninguna normalización, sin embargo se puede afirmar que su construcción sigue las pautas siguientes:

- Identificación de los conceptos importantes de un dominio.
- Clasificación de los más generales a los más específicos.
- Puesta en relación del conjunto.

Los conceptos pueden representar entidades concretas o abstractas, acontecimientos y cosas, a la vez que pueden distinguir entre procesos, procedimientos y productos. Para la descripción de los conceptos tampoco existe ninguna norma o control. Su definición y valor depende íntegramente de su creador. En cuanto a la representación de la semántica de los enlaces, también varía según los autores. Las etiquetas de los enlaces pueden ser verbos o preposiciones, verbos o nombres, verbos o conectores lógicos y en algunos casos los enlaces no vienen etiquetados. En cualquier caso, es importante subrayar la importancia que siempre asume el verbo como encargado de la descripción de los procesos.

En cuanto a la tipología de las relaciones que pueden unir a los nodos en un mapa conceptual, se pueden distinguir los esfuerzos hechos por numerosos investigadores, entre las que pueden citarse las que se emplean con más frecuencia en la mayoría de los dominios del conocimiento:

- Los de directriz (analogía, comentario).
- Relaciones físicas (sobre, debajo, cerca).
- Relaciones temporales (antes, después).
- Relaciones lógicas (causa- efecto).

Finalmente, los mapas conceptuales pueden responder a una organización jerárquica o no jerárquica, con enlaces únicos o múltiples, dispuestos en forma de la tela de araña o de círculos concéntricos, etcétera. Como ejemplo de su uso, se puso en práctica el diseño de un mapa conceptual orientado a la comprensión, de forma general, de los componentes y tecnologías que integran el modelo de la web semántica (fig. 1). El diseño se realizó con el software IMHC Camp Tool, un kit de herramientas orientadas al diseño, mantenimiento e intercambio de mapas conceptuales, desarrollado por el Institute for Human and Machine Cognition (IHMC) de la Florida.

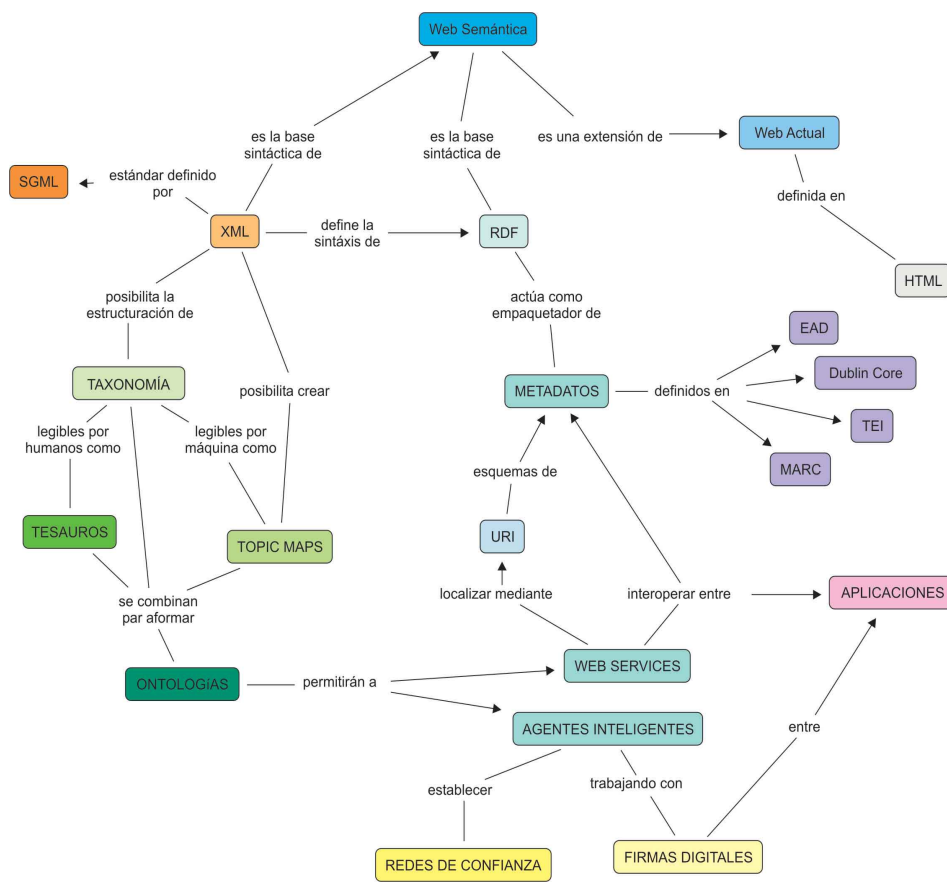


Fig. 1 Mapa conceptual de la web semántica.

Los tesauros

Sobre los orígenes del tesauro, *Norman Roberts*, relaciona de algún modo la aparición de este sistema de organización del conocimiento con tres factores:

- La llamada explosión de la información que dio lugar a la desactualización de los sistemas tradicionales.
- La necesidad de mecanizar los procesos técnicos.

- La teoría de la indización post-coordinada del bibliotecario estadounidense *Mortimer Taube* . 9

El tesauro es un sistema de organización del conocimiento, cuyos términos de indización se estructuran en un sistema de relaciones explícitas que consta de un vocabulario de indización, conformado por descriptores e identificadores. Ambos son términos aceptados en la indización para describir el contenido de los documentos en los catálogos de la biblioteca, bibliografías u otros documentos secundarios; incluyen, a su vez no descriptores que son sinónimos o *cuasi* -sinónimos de los descriptores. Su función es remitir desde un término no autorizado al término equivalencia, cuyos descriptores, identificadores y no descriptores forman el vocabulario de entrada del tesauro. En un tesauro, se hacen explícitos tres tipos de relaciones:

- Relaciones de equivalencia.

- Relaciones de jerarquía.

- Relaciones de asociación.

Un tesauro debe mostrar las tres relaciones; cuando se omiten, se trata de una lista de descriptores, pero no de un tesauro. La sintaxis del tesauro es poscoordinada, es decir, los términos se almacenan en el registro bibliográfico sin seguir un orden de cita predefinido y se coordinan en el momento de la recuperación. Desde el punto de vista de los tesauros, es muy interesante la propuesta que hacen *Corcho, Fernández López y Gómez Pérez* , al mencionar, las ontologías de peso ligero, que incluyen los conceptos, las taxonomías de los conceptos, las relaciones entre conceptos y las propiedades que describen los conceptos. 10 De esta propuesta, puede deducirse claramente su vinculación con los tesauros. Por otra parte, se encuentran las ontologías de gran peso y que superan el ámbito de los tesauros al añadir axiomas.

“Desde hace algunos años, diversos investigadores han defendido la necesidad tanto de cambios estructurales como de adaptaciones tecnológicas por parte de los tesauros”. 11 Desde el punto de vista estructural, cabe subrayar la preponderancia de las relaciones asociativas en detrimento de las jerárquicas, que desaparecerán una vez construido el vocabulario, porque la jerarquía pierde su importancia en favor de las construcciones asociativas, aun cuando mantiene funciones de ayuda a la construcción.

Desde la perspectiva tecnológica, es relevante el creciente interés por su conexión con la recuperación de información, así como otras propuestas como las que hacen referencia a la construcción automática de tesauros, en sus diversas formas. En lo que respecta a su traslación tecnológica, existen nuevos desarrollos de lenguajes documentales en el web, porque los tesauros en línea actuales destacan por su potencialidad hipertextual, empleada para convertir las relaciones entre términos en enlaces, cuestión que también se refleja en los lenguajes epistemográficos. Frente a esta ventaja, los tesauros en línea tienen un alto costo de mantenimiento, en especial en lo que se refiere a la actualización de los enlaces, y requieren de operaciones complejas derivadas de la automatización, pero no son dinámicos ni adecuados

para el tratamiento de objetos, lo que sí parece resolverse con los Topic Map y las ontologías. En el caso de los Topic Maps, lo más significativo es que no establece un número concreto y cerrado de tipos posibles de relación, sino que deja abierta la posibilidad de establecer un número potencialmente infinito de tipos de relaciones.

Berners-Lee, Hendler y Lassila, cuando aluden a la idea de la web semántica, cuyo objetivo es proporcionar un lenguaje que exprese los datos y las reglas para el razonamiento acerca de aquellos, y aportar reglas para que cualquier sistema de representación de conocimiento pueda exportarse al web, señalan a las ontologías como un instrumento útil, porque definen formalmente la relación entre términos, para lo que es necesario contar con una taxonomía y con un conjunto de reglas de inferencia. 5

Las ontologías

El crecimiento de la información en el web ha generado nuevas perspectivas para el diseño de clasificaciones, tesauros y ontologías. Estos lenguajes colaboran en la descripción de los diferentes recursos de información y en su posterior recuperación, lo que también pudiese mejorar la recuperación de información en el web, en términos de efectividad, rapidez y facilidad de acceso a la información. Para la Metafísica, como rama de la Filosofía cuyo estudio se centra en la naturaleza de la realidad última, las ontologías describen los rasgos más generales de esta, como subdivisión que analiza los tipos fundamentales de entidades que componen el Universo, y en la metafísica propiamente dicha, describe los rasgos más generales de la realidad. Sin entrar en la validez de su significado, constituyen un tipo de instrumento que permite la representación del conocimiento en un área determinada en clara conexión con su recuperación en entornos informáticos.

Las ontologías se utilizaron tradicionalmente como modelo de representación de conocimiento en la inteligencia artificial. Según la definición ofrecida por *Gruber* y posteriormente extendida por *Studer, Benjamins y Fensel*, una ontología es “una especificación explícita y formal de una conceptualización”. 12 Más concretamente, una ontología estará formada por una taxonomía relacional de conceptos y por un conjunto de axiomas o reglas de inferencia mediante los cuales se podrá inferir un nuevo conocimiento.

En la literatura especializada, la palabra ontología apunta al mismo tiempo a cosas que, aún estando relacionadas en muchos sentidos, presentan diferencias evidentes. Sobre esta base *Neches*, delimita la ontología desde el punto de vista descriptivo como “un instrumento que define los términos básicos y relaciones a partir del vocabulario de un área de conocimiento, así como las reglas de combinación de estos términos y relaciones para definir extensiones a un vocabulario”. 13 El paralelismo con los tesauros es evidente, especialmente en lo que concierne a la delimitación de los términos que deben formar parte de una construcción de representación del conocimiento en lo que se refiere a las relaciones que se establecen. También pueden identificarse definiciones basadas en los procesos que se han de seguir para la construcción de cualquier ontología, construcción que es considerada desde muchas perspectivas, incluso de hasta quienes incluyen en esta categoría, las clasificaciones utilizadas para la confección de los directorios web como los de Yahoo y Lycos.

Objetivos, tipología y estructura

Noy y McGuinness, plantean que las ontologías tienen como objetivos principales los siguientes: 14

- Compartir la comprensión común de la estructura de información entre personas o agentes de software, lo que debe revertirse de forma positiva y casi necesaria en la extracción y recuperación de información, en páginas web, de contenidos conectados temáticamente.
- Permitir la reutilización del conocimiento perteneciente a un dominio. Por ejemplo, a la hora de iniciar la elaboración de una ontología.
- Permite hacer explícitos los supuestos de un dominio. Esta aseveración puede conducir a conclusiones muy interesantes para la representación del conocimiento más allá de consideraciones técnicas, operativas e informáticas.
- Separa el conocimiento de un dominio del conocimiento que se puede denominar operacional. Con esto se alude a que, en ocasiones, el conocimiento que se está representando se puede implicar en diferentes áreas al pertenecer más a un conocimiento relacionado con procesos.
- Hace posible analizar el conocimiento de un campo, por ejemplo en lo que se refiere al estudio de los términos y relaciones que lo configuran ya sea formalmente o no.

En el contexto de uso actual de las ontologías, como son los portales web, las colecciones multimedia, el diseño de documentos web, los agentes inteligentes, el comercio electrónico, la gestión de la imagen audiovisual, entre otras, pueden contribuir a una mejora en la comunicación con la reducción de la confusión terminológica y conceptual en un único marco de trabajo, así como a su interoperabilidad. En efecto, las ontologías, en principio, han de potenciar el intercambio de datos en contextos informáticos y digitales en entornos heterogéneos y distribuidos gracias a los fundamentos semánticos que se encuentran en ellas.

Según *Guarino* existen ontologías de nivel más alto: 15

- Las de dominios: Están destinadas a describir todos los conceptos generales como el espacio, el tiempo, la materia, el objeto, el hecho, la acción, etc., mediante la clasificación facética.
- Las de tareas: Describen actividades, lo que puede resultar útil en las organizaciones.
- Las de aplicaciones: Las ontologías de aplicaciones describen los conceptos conforme a un campo determinado o tareas concretas, que resultan, en muchas ocasiones, especializaciones de diversas ontologías.

Para *Guerrero y Lozano*, las ontologías más cercanas al dominio de la Documentación serían las ontologías terminológicas porque especifican los términos empleados para representar un conocimiento, y las de Información, al delimitar la estructura de almacenamiento de bases de datos. 16 Su trascendencia documental radica en que cada término y cada relación entre éstos, se define formalmente. Lo que marca la diferencia con los actuales tesauros es la presencia de una mayor variedad de relaciones entre los conceptos, procedentes además del modelo conceptual existente en el dominio formalizado. Asimismo, su función más importante es la de almacenar conocimiento de forma que pueda utilizarse por sistemas automáticos capaces de realizar deducciones a partir de la variedad de relaciones entre conceptos. En cuanto a su estructura, en términos generales, se puede decir que una ontología es una forma más de describir formalmente el conocimiento de un dominio y que se compone de las siguientes partes:

- Clases y subclases: se definen como el conjunto de características que presentan los objetos, especifican relaciones de ubicación en una jerarquía determinada, así como relaciones semánticas entre entidades de información.
- Slots: en ocasiones llamados roles o propiedades, delimitan las propiedades y características de cada concepto, describen varios rasgos y atributos. Ellos ayudan a definir las características de las clases. Pueden ser intrínsecos, extrínsecos, partes o bien relaciones de miembros individuales de las clases y otros ítems.
- Facetas: también llamadas restricciones de roles, describen cosas como los tipos de valores, los valores permitidos, el número de valores y cualquier otra característica que un slot puede tomar. Entre las propiedades más comunes de las facetas están la cardinalidad -permiten definir cuántos valores puede tener un spot-, el tipo de valor del slot (string, número, operadores booleanos), dominio y rango de un slot, etcétera.

También se podrían incluir en esta enumeración básica, formas de cálculo como las funciones y los axiomas, que son teoremas sobre las relaciones que deben cumplir los elementos de una ontología.

Finalmente, destacan las instancias, que en realidad son objetos de una clase.

Métodos, herramientas y lenguajes de codificación

Son muchos los métodos de creación de ontologías propuestos al amparo de diferentes grupos de investigación. Se pueden destacar los métodos de *Uschol and King's*, el utilizado en el proyecto Rodas, el método Methontology, el Onto-Knowledge, etcétera. Así, en ocasiones, se acude a la investigación de las propiedades de los conceptos o de sus relaciones, al uso de esquemas de conceptos semánticos, al uso del modelo de entidad-relación para crear conceptos semánticos con la idea de agrupar entidades y relaciones o al trabajo con estructuras conceptuales asimilables a las redes jerárquicas de trabajo.

Ding y Foo realizan un repaso por los diferentes métodos de trabajo empleados para el diseño de ontologías y establecen los siguientes: 17-18

- *Datos-fuente* : vocabularios controlados, *corpus* de sentencias, extracción en texto libre, preguntas de usuarios, etcétera.
- *Métodos para la extracción de conceptos* : contempla las diferentes técnicas empleadas en la extracción de información: análisis sintáctico, procesamiento del lenguaje natural, implicación humana, etc.).
- *Métodos para la extracción de relaciones* : se aplican normalmente de forma automática a partir diversos algoritmos aunque en ocasiones se aplica de forma manual.
- *Reutilización de ontologías* : suele ser habitual utilizar como base otros instrumentos terminológicos.
- *Representación de la ontología* : que va desde la estructura jerárquica, pasando por la lógica de descripción, hasta los grafos conceptuales y el XML.
- *Herramienta o sistema asociados* : donde se puede observar que no siempre hay programas informáticos inmiscuidos en los proyectos orientados a ontologías.

En general, se puede afirmar que la elaboración y construcción de una ontología debe tener en cuenta su relación con la arquitectura del sistema de información en el que está inmersa sin olvidar la importancia que se deriva de la formular teorías de conocimiento sobre un dominio determinado. Y todo esto sin contar con las diferentes áreas como la evaluación, el aprendizaje, la reingeniería, etcétera. También son variadas las herramientas relacionadas con las ontologías. Desde las iniciales *Ontolingua Server* , *Ontosaurus* a las más recientes como *Protégé 2000* , *WebODE* , y *Ontoedit* .

Las herramientas para la elaboración de las ontologías se subdividen en: 19

- 1) Herramientas de desarrollo de ontologías: este grupo incluye las herramientas que sirven para la construcción de nuevas ontologías o bien para la reutilización de las existentes. Destacan entre sus funcionalidades la edición y la consulta, así como la exportación e importación de ontologías, la visualización en diversos formatos gráficos, etcétera.
- 2) Herramientas de fusión e integración de ontologías: pretenden solucionar el problema de la combinación y la integración de diversas ontologías de un mismo dominio, lo que ocurre cuando se unen dos organizaciones diferenciadas, o cuando se pretende obtener una ontología de calidad, a partir de las existentes.
- 3) Herramientas de evaluación de ontologías: aparecen como instrumentos de apoyo que deben asegurar que tanto las ontologías como las tecnologías relacionadas tengan un nivel mínimo de calidad. Para el

futuro, este esfuerzo puede también conducir a las certificaciones estandarizadas.

4) Herramientas basadas de la anotación: estas herramientas se han diseñado para permitir a los usuarios insertar informaciones y datos. La mayoría de estas herramientas han aparecido recientemente, junto con la aparición de la idea de la web semántica.

5) Herramientas de almacenamiento y preguntas: son instrumentos que se han creado para permitir utilizar con facilidad las ontologías. La clave está en el intento de que el web se convierta en una auténtica plataforma para transmitir conocimiento.

6) Herramientas de aprendizaje: se utilizan semiautomáticamente para construir ontologías a partir del lenguaje natural.

Tesauros y ontologías: diferencias y convergencias.

Son varios los autores que han profundizado en las diferencias existentes entre las ontologías y los tesauros. Por ejemplo, *Qin y Paling* entienden que las primeras son superiores a los segundos por varias razones: en primer lugar, presentan un nivel más alto de concepción y de descripción del vocabulario. Además, las ontologías se caracterizan por un desarrollo semántico más profundo para las relaciones del tipo clase/subclase y para las relaciones cruzadas, lo que supone la ampliación de éstas y un mayor cuidado en su descripción, por supuesto explícita. En tercer lugar, destacan el uso de la lógica empleada en la descripción de situaciones. Y por último, enfatizan en la reusabilidad de las ontologías y en la posibilidad del trabajo en sistemas heterogéneos, al describir formalmente objetos en el mundo, sus propiedades y sus relaciones. 20

En esta misma línea de pensamiento, *Ding y Foo* creen que las diferencias fundamentales entre una ontología y un vocabulario de representación convencional como los tesauros se sitúan en el nivel de abstracción, en las relaciones entre conceptos, en la capacidad para que sean comprensibles para las máquinas y, lo más importante, en la expresividad que pueden proporcionar: 17-18

Una ontología puede estar elaborada de acuerdo con diferentes requerimientos y, al mismo tiempo, puede funcionar como un esquema de base de datos y como una auténtica base de conocimiento para definir varias tareas o aplicaciones.

Una ontología potencia la comunicación entre los humanos y las computadoras mientras que un vocabulario convencional en el mundo de lo que ellos llaman Library Science sólo permite la comunicación entre seres humanos.

Una ontología promueve la normalización y reutilización de la representación de la información mediante la identificación del conocimiento común y compartido.

Las ontologías añaden valor a los tesauros tradicionales por medio de una semántica más profunda, así como desde un prisma conceptual, relacional e informático. De hecho, una mayor profundidad semántica puede implicar niveles más profundos de jerarquía, enriquecidas relaciones entre clases y conceptos, así como la capacidad de formular reglas de inferencia.

No obstante, también se pueden encontrar puntos de encuentro: la creciente complementariedad entre el lenguaje natural y los lenguajes controlados y la disminución en las diferencias entre los distintos lenguajes documentales. Se afirma que las ontologías son instrumentos claramente conectados con los agentes expertos en su objetivo de filtrar información y que, en muchos casos, pueden derivar en tesauros.

Así mismo, *Qin* y *Paling* afirman que las ontologías son instrumentos adecuados para la definición de un vocabulario de representación y coinciden con los tesauros y con las clasificaciones en su capacidad para representar el contenido de un documento por medio de la abstracción y de las relaciones entre conceptos. 20

Para ayudar a comprender qué es una ontología y sus diferencias con los tesauros, muchos especialistas dedicados a la ingeniería del conocimiento suelen hacer una distinción entre dos aspectos muy importantes: términos y conceptos. Tales aspectos establecen diferencias entre los tesauros y ontologías mediante el “*Triángulo del significado*”, que intenta mostrar de forma abreviada los tres componentes del significado de los lenguajes naturales como pueden ser el inglés y el español (fig. 2). Este triángulo básico representa de forma de forma analítica las relaciones existentes entre el significado, el significante y el referente. El significante simboliza un significado, el significado se refiere a un referente, y el referente representa, de forma convencional, al referente. El primer componente, ubicado en el vértice izquierdo del triángulo, alude al término representado mediante símbolos o palabras y las reglas para combinarlas mediante frases y oraciones (sintaxis). Dicho componente, por si solo, no adquiere significado hasta asociarse con otros componentes como los conceptos y sus referentes en el mundo real. 21

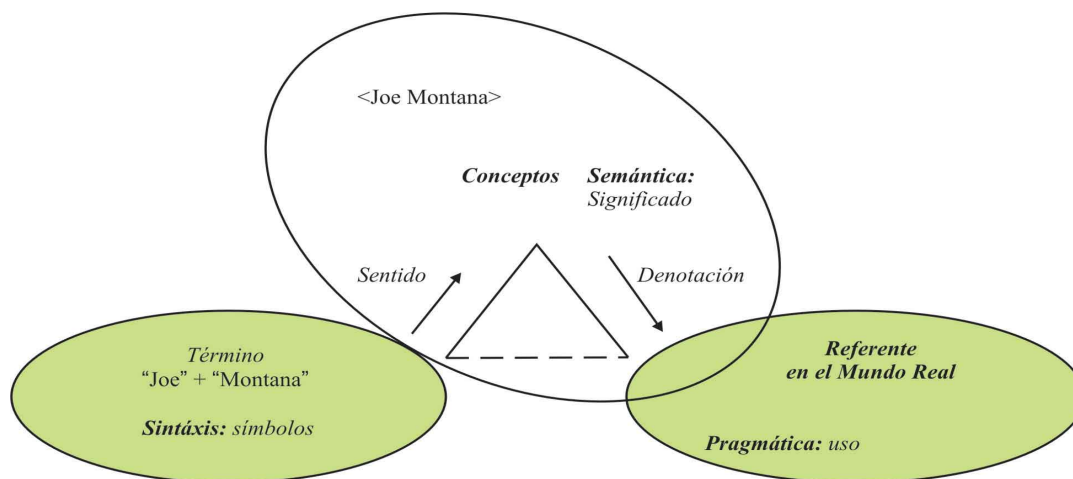


Fig. 2. Triángulo del significado

Los tesauros generalmente operan con el lado izquierdo del triángulo (términos y conceptos), mientras las ontologías lo hacen más en el lado derecho (los conceptos y sus referentes), porque tratan de capturar y representar el significado de un dominio del conocimiento en particular y simularlo explícitamente, siguiendo modelos mentales mediante el razonamiento inferencial.

Representación de las ontologías

Las ontologías definen términos y conceptos utilizados para describir y representar un área de conocimiento. En cuanto a los niveles de representación, es necesario realizar distinciones en cuanto a los niveles de representación, porque pueden considerarse como lenguajes o vocabularios sintácticos acompañados de semántica. Además, las ontologías forman parte del contenido en cualquier entorno de información distribuida o no, y como parte de este solamente pueden expresarse mediante lenguajes de contenidos generalmente denominados “*Lenguajes de representación del conocimiento*”. Según *Daconta, Obrst y Smith* existen tres niveles de representación para las ontologías: 22

- Nivel para la representación de conocimiento (*Metanivel*).
- Nivel para la representación de conceptos (*Nivel de objetos*).
- Nivel para la representación de instancias.

En el primer nivel o metanivel de representación de conocimiento, se definen las construcciones que se utilizarán por las ontologías mediante un lenguaje formal. Estas construcciones acuden a la noción de clases, relaciones, propiedades y axiomas. Los lenguajes de representación del conocimiento, a su vez, pueden dividirse en dos grupos:

- Lenguajes que anteceden al concepto web semántica.
- Lenguajes para la representación del conocimiento en la web semántica.

En el segundo nivel, las ontologías se definen mediante el uso de las construcciones del primer nivel que son las ideas básicas que se intentan formalizar. Los conceptos pueden ser clases de objetos, métodos, procesos de razonamiento, etcétera. Asimismo, el tercer nivel es una extensión del segundo en el que las instancias se utilizan para representar objetos determinados de un concepto mediante bases de conocimiento.

Redes semánticas, representaciones basadas en frames y lógica descriptiva

La representación del conocimiento es una rama de la inteligencia artificial que estudia el diseño e implementación de sistemas y lenguajes para representar el conocimiento sobre el mundo real, con la capacidad adicional de utilizar este conocimiento para el razonamiento inteligente y la eficiencia de los sistemas computacionales. De forma general, los lenguajes de representación del conocimiento con limitaciones de expresividad no pueden emplearse directamente para el diseño de métodos o algoritmos de razonamiento automático que requieren mayor complejidad expresiva. De la representación del conocimiento, se deriva, a su vez, de la noción de redes semánticas, desarrollada durante la décadas de los años 60' y 70' del siglo XX, porque la primera elaboración conceptual de las redes semánticas fue descrita por *Marvin Minsky* en su libro “*Semantic Information Processing*”, donde explica los avances obtenidos en el campo de la inteligencia artificial en el MIT (*Massachusetts Institute of Technology*) en materia de representación del conocimiento y describe “una estructura de red -típicamente representada con grafos acíclicos- para la expresión de semántica o una representación nodo-enlace de construcciones basadas en el conocimiento (conceptos y sus relaciones)”. 23

De ahí que lo que hoy se conoce como mapas del conocimiento probablemente pudieran tener sus bases teóricas en las redes semánticas; el problema con las redes semánticas es que ellas eran estructuras heterogéneas, cuya representación con sus dependencias semánticas, era sólo realizada por los investigadores y los sistemas que las utilizaban sin extensión a otros entornos o aplicaciones computacionales. Los métodos de razonamiento de las redes semánticas se basaron en estrategias particulares de implementación, no en un lenguaje formal.

En un principio, el desarrollo de las redes semánticas no se reconoció como un lenguaje para la representación del conocimiento. “Entre finales de la década de los años 70' y principio de los 80' , la comunidad para la representación del conocimiento comenzó a formalizar coherentemente la noción de red semántica como un lenguaje para la representación del conocimiento basada en la lógica”. 24 La primera formalización de una red semántica basada en la lógica fue el lenguaje basado en frames KL-ONE. Un lenguaje basado en la lógica utiliza frames (marcos), que no son más que simples estructuras de datos para la representación de conceptos (clases) y sus relaciones, concebidos dentro del paradigma de la modelación orientada a objeto, porque el conocimiento de un dominio está volcado sobre entidades primarias -representadas como clases e instancias de estas- del dominio y cuyas propiedades se consideran estándares para modelos de objetos y ontología como: herencia múltiple, valores predefinidos y las facetas, las cuales constituyen aspectos del conocimiento o meta-datos asociados a los frames.

Lenguajes como: DAML+ OIL y OWL, implícita o explícitamente, permiten realizar distinciones para lograr la transparencia comunicativa entre usuario y lenguaje, un elemento importante a considerar a la hora de la interacción hombre-máquina. Las redes semánticas evolucionan hacia sistemas basados en frames potenciado por la lógica descriptiva, cuya descripción es una expresión en lenguaje formal que define un grupo de instancias. Además, es un lenguaje para la representación del conocimiento que consta de sintaxis y semántica, la primera para construir descripciones y la segunda para definir el significado de cada descripción. La lógica descriptiva es llamada en ocasiones lógica terminológica, lógica de clasificación o lógica de concepto basada en formalismos declarativos (axiomas) para la definición conceptual en múltiples taxonomías.

Ontologías y mapeo semántico

Uno de los factores imprescindibles en la comprensión de las ontologías es su representación en un espacio conceptual: *mapeo semántico*. Esta definición corresponde a la confección de diagramas o maquetas visuales que permitan resolver problemas de comprensión en términos de significado entre sistemas, aplicaciones, bases de datos y colecciones documentales. El problema semántico existe dentro y fuera de las ontologías, es decir, en el primer caso el foco de atención se centra sobre el contexto de uso de una ontología específicamente y en el segundo caso en la equivalencia semántica de diferentes conceptos y relaciones entre dos o más ontologías del mismo dominio. Estas ontologías pueden o no referirse a elementos comunes de dicho dominio, debido a que dos ontologías que requieran ser mapeadas de forma conjunta son similares, pero tratan elementos diferentes con relación al mismo dominio. Por otra parte, puede darse el caso de que sea necesario mapear una parte de la(s) ontologías hacia otros estándares de representación como: taxonomías, tesauros, mapas conceptuales, Topics Maps y ontologías más genéricas. En cualquiera de los casos, es necesario evitar la pérdida de significado en el momento de la representación estructural.

Un ejemplo de ello, sería mapear una ontología hacia una taxonomía perteneciente al área del comercio electrónico, específicamente una parte de la taxonomía de la UNSPC (*United Nations Standard Products and Services Code* - <http://www.unspsc.org/>), (fig.3). En la parte derecha de la figura, aparece referenciada una ontología con sus relaciones bien definidas y en la parte izquierda, se muestra la taxonomía empleada por una aplicación de comercio electrónico con menor grado de definición de sus relaciones. En la práctica, se pueden mantener mapas entre ontologías y, como en el caso anterior, entre ontologías y taxonomías, porque cada sistema puede gestionarse por separado en la organización.

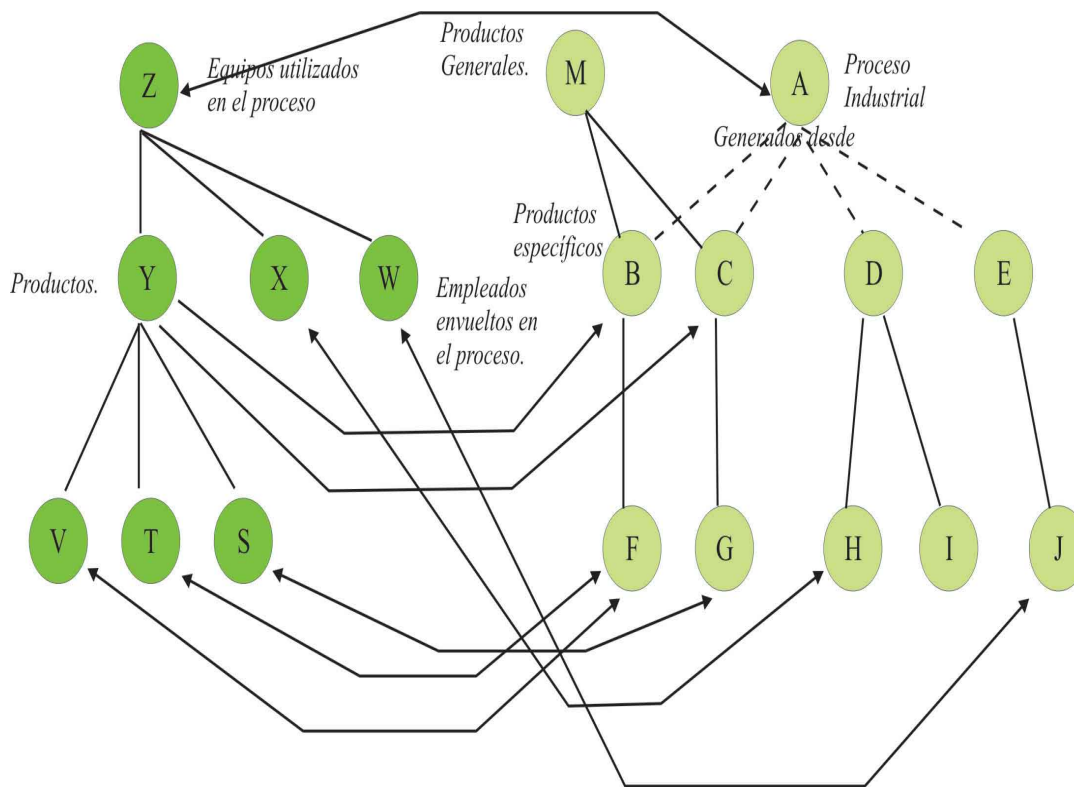


Fig. 3. Mapeo de una taxonomía a una ontología

Determinar la equivalencia semántica entre conceptos en las ontologías es difícil y requiere del conocimiento humano generalmente, aunque las actuales herramientas para la gestión de ontologías proporcionan algún soporte automático para el razonamiento e inferencia de contenido en y entre estas. Aunque las etiquetas que identifican a los conceptos pueden ser las mismas o completamente diferentes en ambas ontologías, no existe garantía de que estos conceptos signifiquen lo mismo o difieran debido a que los términos (palabras) encierran muy poco significado y no pueden emplearse por separado para proporcionar a los sistemas y aplicaciones de software de equivalencia e identidad semántica.

Muchos lenguajes de codificación de ontologías, así como herramientas que le ofrecen soporte, poseen facilidades para la definición de mapas entre ontologías. El mecanismo más simple es incluir o importar declaraciones -en términos de objeto en el paradigma orientado a objeto y el modelo entidad/relación -, porque consiste en insertar una ontología dentro de otra y todos los conceptos y relaciones de la ontología importada estarán disponibles en la nueva integración. No obstante, los nuevos conceptos y relaciones no están rectificadas semántica mente. La fusión de las nuevas y viejas relaciones y conceptos posibilitan mantener las ontologías con sus semejanzas y diferencias, pero conectadas. Aplicaciones como Protégé y Jena, soportan mejor el proceso de fusión y mapeo de ontologías. Los lenguajes concebidos para la web semántica como DAML+ OIL y OWL presentan utilidades para importar ontologías, así como para la declaración de conceptos (representados como clases) e instancias semántica mente iguales, lo cual constituye un nuevo paso hacia el logro de una mayor calidad en la representación de ontologías mediante mapas semánticos, utilizando como soporte la web mediante lenguajes de sintaxis y marcado semántico más expresivos.

Lenguaje de marca: concepto y evolución

Desde hace varios años, los denominados lenguajes de marca han adquirido un lugar muy importante en el procesamiento y recuperación de la información, sobre todo, desde que los documentos electrónicos, alcanzaron gran importancia. Muchos son los lenguajes de marca, pero se destacan SGML y XML, que tienen diferentes desarrollos que pretenden resolver problemas concretos. SGML (*Standard Generalized Markup Language*) comenzó a desarrollarse por IBM en 1969, y se denominó en un primer momento GML (*Generalized Markup Language*).

GML fue el resultado de un esfuerzo por encontrar una vía para conformar la documentación básica de la empresa en un formato electrónico transferible y gestionable por máquina, cuya idea se extendió hasta llegar a convertirse en la norma SGML durante una reunión de la *American National Standard Institute* en 1978. Se consideró como una norma por la ISO en 1986, adoptada como ISO 8879, y pasó a formar parte de un conjunto más amplio bajo el título “*Information Processing- Text and Office Systems*”. 25

SGML se basa en el concepto de *marca*, que originalmente describía una anotación u otro símbolo para indicar a un maquetador de imprenta sobre cómo diseñar una determinada página: tamaño de fuente, márgenes, tipo de fuente, entre otros. El manual de estilo “*The Chicago Manual of Style*”, define marcado como: “*el proceso de marcar un documento manuscrito, cómo se deben utilizar los tipos de letras, los tamaños, espacios de letras, etcétera*”. 26 Un lenguaje de marcado designa a un conjunto de convenciones de marcaje que se utilizan conjuntamente para codificar textos. Un lenguaje de esta naturaleza debe especificar:

- 1) Qué marcas se admiten,
- 2) Cuáles son indispensables.
- 3) Cómo se distinguen las marcas del texto.
- 4) Qué significa cada una de estas.

Según *Van Herwijnen*, SGML proporciona los medios para realizar las tres primeras especificaciones; para la última se necesita concretar una serie de directrices, porque se refiere a la capacidad de codificación del significado de la información. El SGML se diseñó para posibilitar representar la información de manera sencilla y que esta pudiera transferirse entre diferentes sistemas en forma compatible. Se enfrentaba así a otros lenguajes de marca propietarios, dependientes de un software concreto o de una plataforma hardware determinada. 27

De esta forma, SGML permite el intercambio de información entre distintas plataformas, soportes físicos, lógicos y entre diferentes sistemas de almacenamiento y presentación (bases de datos, edición

electrónica) con independencia de su grado de complejidad. La posibilidad de separar el contenido de su formato de presentación permite organizar de manera muy flexible la información, porque su estructuración, actualización, selección, combinación o presentación podrán realizarse según convenga en cada oportunidad. No obstante SGML, aunque conocido por la utilización generalizada de etiquetas descriptivas, no constituye un conjunto predefinido de etiquetas, ni una sintaxis de etiquetado.

Es en realidad un metalenguaje que permite el diseño y control de un conjunto de etiquetas y de una sintaxis a la medida del usuario o de la aplicación. Su adaptación a las necesidades de uso se concreta en la *Declaración de Tipo de Documento* (*Document Type Definition* o *DTD*). La DTD define los tipos de elementos, atributos y entidades permitidas, y expresa algunas limitaciones para combinarlos. Crear una definición de tipo de documento o DTD es como crear nuestro propio lenguaje de marcado para una aplicación específica. Por lo tanto, HTML es una DTD, es decir, dentro de las múltiples posibilidades de SGML, la DTD de HTML describe un conjunto concreto de etiquetas de hipertexto y una sintaxis para utilizarlas en Internet.

Aplicación de SGML

Además de HTML para la web, existen muchas otras aplicaciones de SGML de gran envergadura y trascendencia. Varios organismos y asociaciones profesionales han establecido directrices propias que se materializan en el desarrollo de DTDs especializadas. La Asociación de Editores Americanos contempla el etiquetado para una extensísima categoría de documentos, muchas de las cuales se han reconocido por las asociaciones de estandarización ANSI/NISO e ISO.

Un caso similar es el desarrollo del Formato de Texto Universal (*UTF*), promovido por el Consejo de Telecomunicaciones de la Prensa Internacional y la Asociación de Editores de Prensa Diaria Americana para el intercambio de noticias entre agencias y clientes, que reemplazó a los anteriores formatos IPTC 7901 y ANPA 1312. Este formato responde a la norma internacional “*Unicode*”, para la codificación de caracteres en computadoras, y su objetivo es asignar a cada posible carácter de cada posible lenguaje un número y nombre único. *Unicode*, básicamente, se representa con tres tipos de codificación, según el número de bits necesarios para identificar cada carácter: UTF-8, UTF-16, UTF-32. Actualmente, coincide con el estándar ISO/IEC 10646. Su utilización más frecuente, UTF-8, es compatible con el juego de caracteres ASCII. El UTF-8 tiene la ventaja de ser parcialmente compatible con los programas anteriores que funcionaban con caracteres de 8 bits.

Por otra parte, el proyecto del Consorcio Internacional para el Etiquetado de Textos (*TEI* o *Text Encoding Initiative*) es una iniciativa que ha partido de diversas asociaciones profesionales en el campo de las humanidades. Se trata de definir líneas generales de actuación para la representación de materiales textuales en formato electrónico. El objetivo de TEI es fomentar el uso de etiquetas rigurosas y productivas para cualquier clase de texto, aunque su aporte más directo se produce en el campo de los textos con valor cultural y científico. Estas recomendaciones se recogen en un compendio conocido como *TEI Guidelines for Electronic Text Encoding and Interchange*.

En el ámbito de la descripción archivística, se han desarrollado proyectos para, a partir de la aplicación

de la ISAD (G) [*International Standard Archival Description (General)*], codificarla posteriormente por medio de un lenguaje de marca como SGML, y el empleo de una DTD que se ha denominado EAD (*Encoded Archival Description*), que permitiría su tratamiento electrónico.

Esquema de metadatos

Actualmente la mayor parte del contenido del web está diseñado por humanos para ser leído por ellos mismos, y conforma una caótica biblioteca de enormes dimensiones desde la que resulta complicado y costoso extraer conocimiento alguno.

Para solucionar esto, numerosas comunidades desarrollan y promueven la web semántica, que tiene entre sus objetivos modificar la forma en que se presenta la información en el web de modo que facilite su procesamiento por parte de las máquinas y de esta forma establecer canales para un factible procesamiento, integración y re-utilización de la información contenida en el web, para apostar así por la extracción de conocimiento de mayor utilidad a los humanos.

La web semántica no sería sino a una extensión de la existente, en que la información adquiere significado mediante el uso de meta-datos para proveer una categorización semántica de su contenido y permitir un razonamiento automatizado sobre la información. De manera formal, puede decirse que un metadato es un dato que se encarga de mantener un registro sobre el significado, contexto o propósito de un objeto de información, de forma de poder descubrir, entender, extraer y administrar dicho objeto. En general, estos registros son de menor envergadura que los objetos que describen y se crean en un formato corto y conciso de forma que puedan intercambiarse con facilidad.

Los metadatos tienen sus raíces en el catálogo, probablemente inventado poco después del comienzo de la historia por parte de los sumerios. A lo largo de los siglos, las tabletas de arcilla evolucionaron hasta listas manuscritas y posteriormente, a catálogos de libros después de la invención de la imprenta. Estos primeros catálogos de libros eran impresos y eran listas ordenadas alfabéticamente sin criterios de clasificación sofisticados.

Un avance importante en cuanto a esquemas de clasificación se desarrolla alrededor del 1900 cuando los catálogos de libros se reemplazan completamente por tarjetas, las que entre otras ventajas pueden actualizarse. En la década de los años 60' los métodos de producción en masa (mediante computadoras) hicieron necesaria la existencia de múltiples copias de los catálogos existentes, surgen masivas colecciones distribuidas de libros y los catálogos de tarjetas no logran satisfacer los nuevos requerimientos. Es necesario entonces desarrollar estándares de codificación, llamados hoy metadatos.

Los primeros meta-datos en ambientes digitales y sus bases se desarrollaron a finales del siglo XX, cuando emergen múltiples estándares de codificación, lenguajes y protocolos que se utilizan en la generación y uso de catálogos. Existen tantos esquemas, modelos o estándares de metadatos, como proyectos de creación de sistemas y servicios de información digital en la web. Se pueden distinguir distintas tipologías de esquemas de metadatos como (*Couceiro Ardés D . Metadatos y metainformación*

[Tesis para optar por el título de Licenciada en Ciencias de la Información y Bibliotecología]. La Habana : Facultad de Comunicación, Universidad de la Habana , 2001):

1) Metadatos de propósito general: Aquellos formatos dirigidos a cualquier tipo de información. El ejemplo por excelencia es el Dublin Core, pero también otros formatos como METS (Metadata Encoding and Transmission Standard), dirigido a describir cualquier “objeto documental”, en el seno de una biblioteca digital. En las bibliotecas digitales y, generalmente en la información distribuida en Internet, las colecciones están conformadas por texto, imagen, audio y vídeo en diversos formatos de codificación, es decir, por documentos entendidos como objetos informáticos u objetos de información, de forma abreviada *DLOs* . El acrónimo DLO, y el concepto aparejado a él (Document Like Object), surgió en el seno del desarrollo del modelo de metadatos del Dublin Core, exactamente en el primer taller del DC en Ohio (Estados Unidos), donde comenzó a utilizarse para diferenciar nociones individuales que constituyen un objeto discreto, digno de una descripción individual por medio de metadatos.

2) Metadatos de propósito específico. Se vinculan a un tipo de información digital o a un dominio temático o comunidad informativa como:

- Formatos para describir información de archivo como EAD (Encoded Archival Description) .
- Modelos de metadatos para la información textual de contenido literario y humanístico, como el TEI-Header en la Text Encoding Initiative.
- Esquemas de metadatos para imágenes, como VRA-Core de la Visual Resources Association y SVG (*Scalable Vector Graphics*) derivado de XML.

Además de todos estos esquemas de metadatos, el *World Wide Web Consortium*- W3C en lo adelante- desarrolló RDF (Resource Description Framework), una infraestructura para la descripción de recursos en el web que más que un mero formato de metainformación es un "metamodelo de metadatos" que permite codificar distintos esquemas de metadatos y además, crear otros vocabularios específicos por medio del Lenguaje para la Descripción de Vocabularios RDF (*RDF Schema*).

MARC (Machine Readable Cataloguing)

El formato MARC se concibió para transmitir datos de un sistema a otro y fue revolucionario al incorporar campos de largo variable. Contiene códigos alfanuméricos de extensión fija que determinan el nombre, extensión y donde empieza cada campo de descripción, junto a campos de control, utilizados para clasificar la información en términos de tiempo y lugar. Los campos de descripción variable son los contienen datos de catalogación tradicionales, y están precedidos por un código definido que va desde 001 a 999, donde, por ejemplo, el código 650 es la materia del recurso.

Desde la creación de MARC se generaron más de veinte estándares nacionales (DenMARC, AZMARC,

CHMARC, UKMARC, CAN/MARC, etcétera) los cuales tienden a armonizarse. El más conocido de ellos es USMARC (*United States MARC*), también llamado LC-MARC que fue desarrollado en 1968 por la Biblioteca del Congreso de Estados Unidos y deriva de MARC. El formato MARC es una estructura de datos, basada en la norma ISO 2709 (*Information and Documentation – Format for Information Exchange*) y de orientación inicialmente bibliográfica (*Couceiro Ardés D . Metadatos y metainformación [Tesis para optar por el título de Licenciada en Ciencias de la Información y Bibliotecología]. La Habana : Facultad de Comunicación, Universidad de La Habana , 2001):*

Para la generación de consultas a lo largo de múltiples catálogos en línea, se integra al estándar ISO 23950 (Z39.50) que data de 1988- momento en que fue aprobado por la NISO (*National Information Standards Organization*)- y que permite a un usuario de un sistema de información buscar y recuperar la información sin saber la sintaxis utilizada por los otros sistemas. Posee un protocolo XML llamado XER y es portable a SQL. Ambos, MARC y Z39.50, se utilizan ampliamente por las entidades bibliotecarias tradicionales y es probable que lo hagan por un tiempo debido al alto costo en el que deben incurrir estas entidades para cambiar de formato, además del poco financiamiento de que disponen para estos fines. En el contexto norteamericano, a medida que se hizo evidente la necesidad de informatizar los depósitos de archivos, comenzó a utilizarse una adaptación del formato USMARC: el llamado *USMARC Archives and Manuscripts Control (MARC AMC)*. 28

Actualmente la descripción archivística en formato MARC- producto de la conjugación de CAN/MARC (de Canadá) y USMARC en 1999- se encuentra integrada en el formato MARC21, que sigue vigente en cuanto formato de descripción de archivos, en combinación sobre todo con la norma de contenido Archives, Personal Papers and Manuscripts (*APPM2*). No obstante, en el año 1997, *Daniel Van Pitti*, creador de la DTD de EAD, mencionó algunas de las razones que hacían desaconsejable el uso de MARC en la descripción archivística: 29

En primer lugar, los registros MARC tienen una longitud máxima de cien mil caracteres, y las descripciones archivísticas con frecuencia superan este límite.

MARC acomoda de manera muy pobre la información estructurada jerárquicamente, como lo es la archivística.

La comunidad de usuarios MARC no fue capaz de reconducir el estado del arte hacia el desarrollo de hardware y software de bajo coste.

Pasado cierto período de tiempo, estas afirmaciones pueden carecer de vigencia por cuanto el desarrollo del software ha abaratado los costos de producción, al tiempo que incorporado técnicas de programación web y de orientación a objetos que minimizan, cuando no eliminan, los problemas de la extensión y la jerarquía.

Sin embargo, el formato MARC ha encontrado nuevos retos que no ha resuelto adecuadamente. A medida que el proceso de codificación se hace más sofisticado, en un entorno dominado por redes cada

vez más globales, por metadatos cada vez más complejos y por necesidades de recuperación que exigen manejabilidad, la sofisticación parece inevitable a los efectos de una precisión, se requieren más códigos, no externos, sino derivados de la propia estructura MARC. Además, en la medida que su estructura se define según una posición es limitado, e intentar dar cabida a todas estas nuevas realidades es un propósito que está conduciendo a la saturación y la inestabilidad.

MARC es un excelente contenedor de datos en la medida que permite fragmentar de manera coherente la información y, en consecuencia, realizar procesos más complejos con ella. Pero para poder sobrevivir, el formato tendrá, en primer lugar, que cambiar de orientación y en segundo lugar, reconocer que no fue diseñado para cualquier propósito y definir sus límites.

TEI (Text Encoding Initiative)

Comenzó por un proyecto de investigación internacional patrocinado por la Association for Computing in the Humanities (*ACH*), la Association for Literary and Linguistic Computing (*ALLC*) y la Association for Computational Linguistics (*ACL*) en 1987; se orientó a la codificación de textos, fundamentalmente literarios, posterior a 1994. *TEI* ha desarrollado esquemas de codificación modular basados en SGML para una amplia variedad de tipos de documentos. Como consecuencia, el número de textos electrónicos en el área de humanidades ha crecido considerablemente y muchas bibliotecas e instituciones universitarias han comenzado a ampliar sus colecciones de textos electrónicos basadas en estas directrices.

El encabezamiento TEI (*TEI Header*) constituye una de las principales contribuciones para la codificación SGML de información bibliográfica. Las directrices *TEI* incluyen una sección especial - (24.3)- sobre los elementos de este encabezamiento y sus relaciones con los registros *Marc* . Estos elementos contienen meta información sobre el texto, con la particularidad de que algunas partes de estos encabezamientos se realizan según las normas de descripción existentes en los catálogos bibliotecarios como *AACR2* , *ISBD (G)* y *USMARC*. Esto no es sorprendente, si se sabe que dicho encabezamiento fue diseñado por el *Committee on Text Documentation* compuesto, entre otros, por archiveros y bibliotecarios.

Dunlop comenta el uso del *TEI Header* en el *British National Corpus Project* para la descripción a nivel de colección. 30 Esta información descriptiva, que es común para muchos textos de este repertorio, puede utilizarse para la generación automática de encabezamientos en una estructura de base de datos relacional. El proyecto *Digital Imaging* de la *Bodleian Library* de Oxford, utiliza el modelo *TEI* para la creación de la base de datos como un esquema de metadatos conectado a las imágenes digitalizadas. La recuperación proporciona una imagen del documento con sus datos bibliográficos asociados.

EAD (Encoded Archival Description)

Comenzó a desarrollarse en el marco del *Berkeley Finding Aids Project*, con SGML como metalenguaje para aplicar la codificación normalizada a los instrumentos de descripción archivística. El resultado de este proyecto fue una DTD que define una clase de documento que, en términos generales, consta de una

página de título opcional, la descripción de una unidad de material archivística y una serie de apéndices opcionales.

Actualmente, el uso de *EAD*, bajo mantenimiento de la Biblioteca del Congreso de Estados Unidos y responsabilidad de la Sociedad Americana de Archiveros, se ha generalizado para las descripciones de documentos de archivo potencialmente conectadas a sus correspondientes imágenes digitales y su difusión por medio de Internet. La *DTD*, actualmente en su versión 1.0 y que posibilita el empleo de XML, fue diseñada para reflejar la jerarquía natural que presenta la organización de los fondos, en conjunción con el orden intelectual que imponen los archiveros con sus prácticas descriptivas. Contiene dos tipos de elementos:

Los que codifican puntos específicos en la definición de partes componentes del instrumento de descripción o el material que describe (descriptivos).

Los que podrían codificar cualquier característica del documento (genéricos).

A un nivel muy básico, un modelo de documento codificado utilizando *EAD*, consta de tres segmentos:

Uno que proporciona información sobre el instrumento de descripción propiamente dicho: su título, compilador, fecha de compilación (*<eadheader>*, que está basado en el *TEI header*).

Un segundo componente que incluye las cuestiones preliminares necesarias para la publicación formal del instrumento de descripción (*<frontmatter>*).

Y un tercero que proporciona la descripción del material archivístico, además de la información contextual y administrativa asociada (*<findaid>*).

Burnard y Light concluyeron que los componentes del encabezamiento *EAD* difieren ligeramente de los de *TEI*. Estos autores consideran la *EAD* como metadatos puros, debido a que los instrumentos de descripción son metadatos. En este sentido, el *EAD header*, que describe el instrumento de descripción, puede considerarse como metadatos. 31

Dublin Core

Desde su aparición en 1995, se orientó a promover un estándar de propósito general, sencillo y descriptivo de los recursos web sobre cualquier materia. Originalmente facilitaría la indización consistente sin necesidad de una gran experiencia en la aplicación de procesos técnicos como la catalogación y clasificación de información. Promovidos concretamente por la iniciativa DCMI (Dublin Core Metadata Initiative) y su conjunto de elementos ISO 15836-2003, constituye una de las infraestructuras operacionales de la web semántica y una de las claves de la interoperabilidad de la información electrónica.

El Dublin Core es un conjunto de quince elementos básicos para describir el contenido web según tres tipos de información:

- Siete elementos denominados de contenido estrictamente (Título, Autor, Claves, Descripción, Fuente, Lengua, Cobertura).
- Elementos con información relativa a la propiedad intelectual del recurso (Creador, Editor, Otros colaboradores, Derechos).
- Elementos relativos a la temporalidad y formato del documento, así como su identificación (Fecha, Tipo de recurso, Formato, Identificador del recurso).

Además de estos elementos básicos (ninguno obligatorio y todos repetibles) existen otros mecanismos que sirven para adaptar el Dublin Core a las necesidades concretas de información y que hacen que este modelo de meta-datos sea aplicable a cualquier proyecto de sistema o servicio de información digital. Estos mecanismos son fundamentalmente (*Couceiro Ardés D . Metadatos y metainformación [Tesis para optar por el título de Licenciada en Ciencias de la Información y Bibliotecología]. La Habana : Facultad de Comunicación, Universidad de La Habana , 2001*):

Perfiles de aplicación , desarrollados para el uso del Dublin Core asociado a distintas disciplinas, como por ejemplo el perfil de aplicación para bibliotecas DC-Lib.

Términos de metadatos , donde se incluyen los nuevos elementos que se incorporan al vocabulario Dublín.

Core , como las matizaciones de elementos existentes, esquemas de codificación (antes conocidos como calificadores) y términos de vocabularios controlados. Todos estos términos de metadatos sirven para adecuar y precisar el valor y la utilidad de la metainformación expresada a través de Dublin Core.

La adaptabilidad de este modelo es un factor determinante para la descripción de recursos web mediante la codificación sintáctica en RDF/ XML como esquema genéricos para el intercambio de metadatos entre sistemas heterogéneos y distribuidos, estándares de descripción de contenido (ontologías, Topic Maps, tesauros, etc.) y toda una gama de protocolos y normas para el intercambio de información, constituyen la base para la segunda generación del web.

XML (EXTENSIBLE MARKUP LANGUAGE)

El desarrollo del web ha planteado nuevos retos para tratar de manejar adecuadamente la gran cantidad de información que por ella circula diariamente. Se plantea además el problema de su falta de estructuración y la complejidad que adquiere, que dificultan los procesos de gestión, mantenimiento y recuperación de información. XML nace en el año 1996- ratificada por el *World Wide Web Consortium* en 1998 mediante la especificación XML 1.0- con el fin de posibilitar el intercambio de documentos

estructurados por medio del web, a la vez que permitía el uso del hipertexto. XML es una versión abreviada de SGML, lo que permite una mayor agilidad en su implementación, pero es necesario considerar que carece de muchas de las características del propio SGML y precisamente por ello no lo sustituye. Los objetivos de XML, según esta recomendación del Consorcio Web, son los siguientes: 32

- Se ha creado para el web.
- Sobre XML se puede desarrollar una gran cantidad de aplicaciones.
- XML debe ser compatible con SGML.
- Debe ser fácil la escritura de los programas que procesen los documentos XML.
- Por compatibilidad, el número de características opcionales debe tender a cero.
- Los documentos XML deben ser legibles por los usuarios y no desvirtuar el contenido.
- El diseño de XML debe ser muy rápido.
- Los documentos XML deben ser fáciles de crear.

XML precisa de otros estándares asociados como, Unicode e ISO/IEC 10646 para el tratamiento de caracteres, Internet RFC 1766 para las marcas de identificación de lenguaje, ISO 639 para los códigos de nombre de lenguaje e ISO 3166 para los códigos de nombre de país.

Los documentos XML están formados por unidades de almacenamiento llamadas entidades que contienen datos procesados o sin procesar. Los datos procesados están formados por caracteres, algunos de los cuales forman datos de carácter, y otros marcas. Las marcas codifican la descripción del esquema de almacenamiento y la estructura lógica del documento, y pueden establecer mecanismos de restricción, tanto al esquema de almacenamiento como a la estructura lógica. Su potencia proviene de la separación que ofrece entre la interfaz de usuario y la estructura de los datos, se centra en la definición de los contenidos. Se separan los datos de la representación y del procesamiento, así permite mostrar y procesar los datos como se desee, en dependencia de las diferentes aplicaciones u hojas de estilo empleadas.

Uno de los aspectos más importantes del XML es que es un conjunto de tecnologías basadas en estándares abiertos, que forman módulos opcionales y que amplían sus posibilidades. Algunos de estos módulos son: 32

DTD (Document Type Definition): Conjunto formal de declaraciones de elementos, atributos y entidades que le indican a un sistema exactamente el tipo de etiquetado que se utiliza en dicho documento.

Esquemas para XML (XML Schema): Si bien las DTD permiten describir documentos, un esquema es algo más restrictivo, más similar a un esquema de bases de datos en que el contenido de los elementos tiene asociado un tipo. Un esquema permitiría a un procesador validar el documento por inconsistencias de una forma más apropiada.

XML Namespaces : Especificación que describe como asociar un “URI”, con cada etiqueta y atributo en un documento XML. El Uniform Resource Identifier es un sistema ideado por el Internet Engineering Task Force (IETF) que pretende ser un sistema global para identificar recursos en el web, sean documentos, imágenes, archivos de programas, correos electrónicos, entre otros. Es un método genérico que combina URL y URN (*Uniform Resource Name*) y que se refiere al subconjunto de URIs que se requiere para identificar un recurso de forma globalmente única y persistente, incluso si el recurso deja de existir. El término más significativo es el de *identificadores* , que no es más que una secuencia de caracteres con sintaxis controlada mediante el cual se reconoce un documento de forma única. 32

CSS : El lenguaje de hojas de estilo.

XSL : Lenguaje avanzado para expresar hojas de estilo, que se basa en XSLT (*eXtensible Stylesheet Language Transformation*) para la transformación de documentos desde el punto de vista del contenido.

DOM : Un conjunto estándar de funciones para realizar llamadas, desde un lenguaje de programación, para manipular ficheros XML.

Xlink : Describe como añadir hiperenlaces a un fichero XML.

XPointer y *Xfragments* : Son sintaxis que indican la forma de apuntar a una parte, dentro de un documento XML.

Estructura de los documentos XML

Definición de tipo de documento y esquemas XML

A partir de que los desarrolladores web puedan crear su propio vocabulario, las organizaciones requieren establecer estándares de estructuras de documentos para que sus miembros puedan comunicarse entre ellos. La especificación de la estructura de un documento XML con independencia de su contenido, permite el control de la consistencia de documentos creados por diferentes autores, cuyo propósito se destina a una Definición de tipo de documentos (*DTD o Document Type Definition*) o un esquema en XML (*XML Schema*).

Una DTD especifica los elementos, atributos, entidades y relaciones permitidos en un documento XML. Puede estar incluida en el mismo documento o en un documento separado que se puede referenciar

mediante un URL. Describe los datos, proporciona la gramática y un vocabulario del lenguaje que se utiliza en un documento XML, así asegura que los autores comprendan la estructura de los datos y los usuarios reciban toda la información que necesitan de forma consistente. Una DTD usa una sintaxis que no está basada en XML para definir elementos, atributos y relaciones mediante un pequeño conjunto de sentencias declarativas, un conjunto de símbolos que ayudarán a definir la estructura de los datos y palabras reservadas para especificar tipos de datos.

Actualmente, se definen DTD para grupos sectoriales con intereses comunes de forma que existan estándares avalados por organizaciones que garanticen que cualquier usuario adopte las mismas normativas. Como ejemplos de estas DTD estándares, se pueden identificar: 32

- MathML (*Mathematical Markup Language*): Para el intercambio de datos matemáticos.
- SMIL (*Synchronized Multimedia Integration Language*): Para el intercambio de presentaciones multimedia.
- RDF (*Resource Description Framework*): Como plataforma contenedora para el intercambio de metadatos.

No obstante, se impone otra forma más eficaz de definición de elementos conocida como esquemas. Se puede definir como una DTD que permite su ampliación mediante un lenguaje de definición de esquemas (*XML Scheme*) y su funcionalidad equivale a una DTD, pero con la diferencia de que está escrito en sintaxis XML. Asegura, por tanto, que los documentos se adhieran a un vocabulario previamente definido para permitir la extensión de la funcionalidad de una DTD con tipos de datos, herencia y reglas de presentación.

Espacios de nombre (Namespaces)

XML fue creado con la finalidad de permitir la interoperabilidad entre aplicaciones y esquemas de metadatos en el que sus creadores pueden diseñar y mantener sus propios vocabularios en esta sintaxis. Se puede producir entonces una gran confusión si diferentes desarrolladores escogiesen los mismos nombres de elementos para representar diferentes entidades; por ello, se introdujeron los espacios de nombres (*namespaces*) en XML para resolver este problema, y posibilitar así el uso de múltiples vocabularios en un mismo documento.

Se puede definir y utilizar entonces un espacio de nombre dentro de un URI para resolver conflictos de nombres entre elementos de un documento XML cuando los elementos se derivan de disímiles fuentes (fig. 4). En este caso, la declaración de *namespace* (*xmlns*) se le asocia el prefijo “edi” con el URI “<http://www.amazon.com/pc.htm>”. El nombre del elemento “precio” utiliza el prefijo para crear un nombre calificado, e indica que el elemento “precio” se deriva del espacio de nombre especificado.

```
< x xmlns: edi:= << http://www.amazon.com/pc.htm >> >  
< edi: precio unidades = << Euro>>> 450.95 </edi: recio>  
< x >
```

Fig.4. Declaración de un espacio de nombre en XML.

Estilo y presentación

Debido a que el objetivo fundamental de XML es separar el contenido de su presentación, cualquier elemento XML puede enlazarse a una hoja de estilo para su visualización, específicamente con una hoja de estilo de cascada (*CSS o Cascade Style Sheet*). Esta especifica el formato de presentación; fuentes, colores, márgenes, etc., en que inicialmente se presentaban los documentos HTML codificados en el navegador web. De ahí que el lenguaje extensible de CSS (*XSL o eXtensible Stylesheet Language*) utilice la sintaxis XML, para proporcionar un conjunto de herramientas de formato mucho más completa que CSS. 32

En los documentos sustentados sobre la base de SGML no es posible codificar la información de presentación en el navegador debido a la variedad de tipos de documentos que se aceptan en el sistema. Por esta razón, estos sistemas necesitan un Lenguaje de Especificación y Semántica de Estilo de Documentos (*DSSSL o Document Style Semantics and Specification Language*) para proporcionar la presentación e información de estilo, que es, a su vez, un estándar basado en SGML que regula las normas de presentación de documentos de marcas para la web. En el caso de los navegadores XML o analizadores sintácticos (*parsers*), aceptan tipos de documentos diferentes y por tanto necesitan un lenguaje de estilo.

Por la complejidad del lenguaje de estilo del SGML, para lograr un uso amplio en el web, se desarrolló XSL como subconjunto de DSSSL con medios adicionales para operar en ambientes basados en HTML y XML. Además, posibilita especificar la presentación de los datos según ciertas condiciones, se acerca un poco a un lenguaje de programación y posibilita la ejecución de bucles, sentencias condicionales, operaciones lógicas, ordenación de datos y utilización de plantillas, entre otras facilidades.

Transformación e integración de datos

Otro de los grandes beneficios de XML es su capacidad para transformar el código desarrollado para una aplicación, y hacerlo corresponder con el código desarrollado por otra. En este sentido, XSL se utiliza

para especificar la presentación de un documento- como las hojas de estilo CSS en HTML-, pero las propiedades adicionales de XSL también posibilitan transformar la estructura de un documento XML mediante la transformación de datos con XSLT (*XSL Transformation*).

En este sentido, otra aplicación importante de XML es el intercambio de datos electrónicos entre dos o más fuentes de datos en el web orientado al procesamiento por computadora. Desarrolladores de software, organizaciones de información y gestores de información podrían automáticamente integrar información de fuentes relacionadas que publican sus datos en XML, también las organizaciones podrían publicar datos sobre sus productos, servicios y clientes potenciales pudiesen comparar y procesar esta información. Es posible también presentar nuevas oportunidades a terceros a partir de la integración, transformación, limpieza y agregado datos en XML. En este contexto, se toma como punto de partida, el enfoque de bases de datos de XML, y se considera que un documento XML es una base de datos y su DTD es el esquema de base de datos. Las aplicaciones de intercambio de datos electrónicos requieren herramientas que soporten tareas de:

- Extracción de datos de extensos documentos codificados en XML.
- Conversión de datos entre bases de datos relacionales orientadas a objetos y XML.
- Transformación de datos de una DTD a otra diferente.
- Integración de datos XML de múltiples fuentes.

La extracción, conversión, transformación e integración de datos son problemas conocidos en el enfoque de bases de datos, cuyas soluciones se basan generalmente en un lenguaje de consulta relacional como SQL (*Structure Query Language*) u orientado a objeto (*Object Query Language*). “Para XML se han propuesto diferentes lenguajes de consulta como XQL, Lore y XML-Query Language”.³³ En el caso de XML-Query Language, presenta las siguientes características:

- Es declaratorio.
- Es relacionalmente completo y, en particular, puede expresar combinaciones.
- Es lo bastante sencillo como para que técnicas de bases de datos conocidas para la optimización de consultas y estimación de costos puedan extenderse a XML-QL.
- Puede extraer datos de documentos XML existentes y construir nuevos documentos XML.

Para una aplicación hipermedia, donde los datos provienen de un conjunto de aplicaciones distinto, los datos pueden almacenarse en un sistema gestor de base de datos relacional, orientado a objeto o documentos XML propiamente dichos que tendrán estructura y orígenes heterogéneos. El trabajo con distintas fuentes de datos es aplicable a la arquitectura de Sistemas Federados, donde se distinguen los

siguientes esquemas: 32

- *Esquema local*: Los datos y la estructura se mantienen como los facilita la aplicación emisora. En este caso, un sistema de base de datos relacional tendrá como esquema un conjunto de tablas interrelacionadas. Si la aplicación facilita XML, el esquema corresponderá con la DTD asociada.

- *Esquema componente*: Representación del esquema local en términos del modelo canónico de datos. En este caso se escoge XML como lenguaje común. Si la aplicación emisora no utiliza este modelo, será necesario un conversor para pasar el esquema local a XML.

Esquema federado: Esquema resultante de la integración de varios esquemas exportados utilizando mecanismos de vistas. “Para el caso de documentos XML, una posibilidad es utilizar XSLT o XML-Query Language”. 34

Progresivamente han aparecido soluciones a diferentes problemas documentales en el web, basadas en XML y RDF (*Resource Description Framework*) para el intercambio de meta-datos.

RDF (RESOURCE DESCRIPTION FRAMEWORK)

En un principio el web se construyó para el uso humano y, a pesar de que todo en ella era legible por máquina, los datos todavía no lo son del todo. Una de las soluciones es el uso de metadatos para describir los datos contenidos en ella mediante empaquetadores genéricos que permitan la integración y comunicación entre distintos esquemas de meta-datos. RDF se creó en agosto de 1997, bajo el auspicio del W3C con la finalidad de crear una infraestructura para la descripción de recursos que proporcione una base para procesar meta-y posibilitar la interoperabilidad entre aplicaciones que intercambian información legible por máquina en el web.

XML provee información semántica como un subproducto de definir la estructura del documento, porque define una estructura en forma de árbol para un documento, de manera que sus hojas contienen la información. Se puede observar entonces que la estructura y la semántica de un documento XML están entrelazadas, porque RDF provee medios para agregar semántica a un documento sin referirse a su estructura, como la DTD recomendada por el W3C. Su objetivo general es definir un mecanismo para describir recursos de forma que no cree dependencia alguna sobre un dominio de aplicación particular, ni defina, a priori, la semántica de algún dominio de aplicación. La definición del mecanismo debe ser neutral con respecto al dominio, sin embargo este debe adecuarse para describir información sobre cualquier dominio. Entre los aspectos más representativos de la semántica funcional de RDF, salidos a la luz en la Recomendación del W3C del 22 febrero de 1999, se encuentran:

- El modelo de datos.

- La sintaxis.

- El esquema

El modelo de datos

El fundamento o base de RDF es un modelo para representar propiedades designadas y valores de propiedades. El modelo RDF se basa en principios perfectamente establecidos de varias comunidades de representación de datos como el *Warwick Framework*. Las propiedades RDF pueden recordar atributos de los recursos y, en este sentido, corresponden con los tradicionales pares de atributo-valor.

Las propiedades RDF representan también la relación entre recursos y por lo tanto, un modelo RDF puede parecer un diagrama entidad-relación. De forma más precisa, los esquemas RDF —que son objetos específicos de la categoría del modelo de datos RDF— son diagramas entidad-relación. En la terminología del diseño orientado a objetos, los recursos corresponden con objetos y las propiedades corresponden con objetos específicos y variables de una categoría.

El modelo de datos de RDF es una forma de sintaxis-neutral para representar expresiones RDF y se usa para evaluar la equivalencia en significado: dos expresiones RDF son equivalentes y sólo si sus representaciones del modelo de datos son las mismas. Esta definición de equivalencia permite algunas variaciones sintácticas en expresiones sin alterar el significado. Así el modelo de datos provee tres tipos de objetos: *recursos*, *propiedades* y *sentencias*: 35

- Recursos: Todas las cosas descritas por expresiones RDF se denominan *recursos*. Un recurso puede ser una página web completa; como el documento HTML ("<http://www.desarrolloweb.com/programacion.php>" por ejemplo). Un recurso puede ser una parte de una página web, es decir, un elemento HTML o XML específico dentro del documento fuente. Un recurso puede ser también una colección completa de páginas de un sitio web completo, así como puede ser también un objeto que no sea directamente accesible vía web como, por ejemplo, un libro impreso. Los recursos se designan siempre por URIs, más los identificadores de anclas opcionales. Cualquier cosa puede tener un URI; la extensibilidad de URIs permite la introducción de identificadores para cualquier entidad imaginable.

- Propiedades: Una propiedad es un aspecto específico, característica, atributo, o relación utilizado para describir un recurso. Cada propiedad tiene un significado específico, define sus valores permitidos, los tipos de recursos que puede describir, y sus relaciones con otras propiedades.

- Sentencias: Un recurso específico junto con una propiedad denominada, más el valor de dicha propiedad para ese recurso es una sentencia *RDF* (*RDF statement*). Estas tres partes individuales de una sentencia se denominan, respectivamente, *sujeto*, *predicado* y *objeto*. El objeto de una sentencia (el valor de la propiedad) puede ser otro recurso o puede ser un *literal*; es decir, un recurso (especificado por un URI) o una cadena simple de caracteres u otros tipos de datos primitivos definidos por XML. En términos RDF, un literal puede comprender en su contenido marcado XML pero no puede valorarse más

por un procesador RDF, porque existen varias restricciones sintácticas en cómo se puede expresar el marcado en literales.

Sintaxis de RDF

El modelo de datos RDF proporciona un marco abstracto y conceptual para definir y utilizar metadatos. Necesita también una sintaxis concreta para crear e intercambiar metadatos. Esta especificación de RDF utiliza el XML codificado como su sintaxis de intercambio. RDF necesita también la facilidad de los espacios de nombres en XML (*XML namespace*) para asociar con precisión cada propiedad con el esquema que define dicha propiedad. Esta especificación define dos tipos de sintaxis XML para codificar una instancia (objeto específico de una categoría) de modelo de datos.

La sintaxis seriada básica: Expresa las capacidades totales del modelo de datos de una forma muy regular.

La sintaxis abreviada: Aunque la sintaxis seriada muestra la estructura de un modelo RDF más claro, normalmente es mejor utilizar una forma XML más compacta. Esto se realiza por medio de la sintaxis abreviada de RDF. Como valor añadido, la sintaxis abreviada permite a los documentos seguir los DTDs de XML bien estructuradas que se interpretan como modelos RDF, e incluyen construcciones adicionales.

El esquema

A pesar de todo, el modelo y la sintaxis no facilitan los mecanismos para la definición de propiedades y relaciones entre predicados y objetos. Para ello, desde el año 2000, se definió una especificación para definir esquemas (*schemas*). Un esquema RDF es un conjunto de informaciones relativas a las clases de recursos que sirve para explicitar relaciones jerárquicas que se establecen entre ellos, o bien para matizar el carácter obligatorio u opcional de las propiedades y otras restricciones como el número de ocurrencias.

En otras palabras, puede pensarse en un esquema como en una especie de diccionario. Un esquema define los términos que se utilizarán en una declaración RDF y le otorgará significados específicos. Con RDF se pueden utilizar una gran variedad de formas de esquema, incluso una forma específica (*RDF Schema*) que tiene algunas características específicas para ayudar a la automatización de tareas con RDF.

Un esquema es una representación donde se documentan o explican las definiciones y restricciones de uso de las propiedades. El mayor problema es generalmente la diversidad de intereses de las distintas comunidades de usuarios del web, que provoca, a su vez, ciertas diferencias terminológicas. Para evitar confusiones entre definiciones independientes y posiblemente conflictivas del mismo término, RDF

utiliza la facilidad de los espacios de nombre (*namespaces*) de XML, que permiten expresar un espacio o esquema inequívoco al consignar un recurso que define la semántica correspondiente al principio de un registro de meta-datos. Los espacios de nombre son simplemente una forma de asociar el uso específico de una palabra en el contexto del diccionario en que se puede encontrar una definición determinada. En RDF, cada predicado utilizado en una declaración debe identificarse con un sólo espacio de nombre o esquema (fig. 5).

```
<xmlns: rdf="http:// www.w3.org/TR/REC-rdf-syntax">  
< xmlns: rdfs="http:// www.w3.org/TR/PR-rdf-schema">  
< xmlns: dc=" http:// www.purl.org/DC">
```

Fig. 5. Codificación en RDF para el registro de metadatos en Dublín Core.

Según la figura anterior, se cualificará el nombre de los elementos y atributos utilizados en el registro RDF de la siguiente forma:

En primer lugar, según el *Resource Description Framework* como mecanismo para expresar el modelo y la sintaxis de metadatos RDF.

La propuesta de recomendación para el esquema.

El modelo, en este caso, *Dublín Core* como vocabulario de designación de tipos de atributos definidos por la comunidad que lo desarrolla.

De esta forma, las declaraciones correspondientes a cada espacio de nombre, con sus respectivos URIs, definen los esquemas correspondientes cuyos prefijos para las propiedades serán: “rdf”, “rdfs” y “dc”. El uso de conjuntos de metadatos que garanticen la interoperatividad entre diversos sistemas de información como puede ser el uso de *RDF* y *Dublín Core* , la utilización de lenguajes de etiquetado menos complejos que el SGML y menos simples que el HTML como es el caso de XML, la aplicación de protocolos pensados para la recuperación de información entre aplicaciones como los Web Services y el empleo de técnicas de recuperación de información para generar servicios determinados, permitirán que este sistema de meta-datos tenga un futuro bastante promisorio por cuanto se fundamenta en cuatro principios según el W3C: 36

- Neutralidad: Por no estar ligado a ningún otro sistema de metadatos.

- Expresividad: Uso de etiquetas de forma intuitiva y facilidad de expresión del contenido.

- Familiaridad: Su base en SGML posibilita que sea asequible para las personas relacionadas con HTML.

- Sencillez en el procesamiento: Por ser texto ASCII.

Por estas razones, “han aparecido en los últimos años gran cantidad de aplicaciones que facilitan el trabajo con este sistema de metadatos”, 36 como: “APIs” - API es la abreviatura de *Application Programming Interface* . Un API no es más que una serie de servicios o funciones que el sistema operativo ofrece al programador. Se refiere a toda interfaz de programación que un sistema operativo o aplicación exporta para su uso por ella misma o por otras aplicaciones. Por citar un ejemplo, se destaca *JENA*.

Se trata de una aplicación experimental diseñada por *Hewlett Packard* en *Java* para manipular modelos *RDF* . *JENA* permite trabajar con contenedores y, además, incluye un analizador sintáctico capaz de procesar información RSS (*RDF Site Summary*). RSS es un vocabulario RDF basado en XML que permite la catalogación de información (noticias y eventos) de manera que sea posible encontrar información precisa adaptada a las preferencias de los usuarios. Los archivos RSS contienen metadatos sobre fuentes de información especificadas por los usuarios cuya función principal es avisar a los usuarios de que los recursos que ellos han seleccionado para formar parte de esa RSS han cambiado sin necesidad de comprobar directamente la página, es decir, notifican de forma automática cualquier cambio que se realice en esos recursos de interés seleccionados, ello lo convierte en un módulo ideal para cualquier proceso automatizado de creación de contenidos destinados al uso en canales de información.

A partir de las descripciones precisas de los recursos de información codificados mediante RDF, un agente inteligente será capaz de recuperar información relevante y precisa de diferentes repositorios de información accesibles en el web. A pesar de lo novedoso que pueda resultar el término web semántica, pueden apreciarse algunas evidencias que no son nuevas para los profesionales de la información:

La necesidad de aplicar esquemas de organización en la descripción de recursos y en la formalización de las necesidades de información de los usuarios del sistema.

El propio hecho de que las descripciones de los recursos sean precisas, hacen posible la discriminación de información relevante.

Una de las iniciativas estrechamente relacionadas con estos problemas, específicamente a la web semántica y que se suele relacionar con el modelo RDF y los sistemas de metadatos para web, son los Topics Maps.

TOPIC MAPS

Un Topic Map tiene como finalidad normalizar los elementos y la notación utilizada para estructurar la información mediante la construcción de una red de enlaces semánticos que relacione diferentes recursos

informativos. Tienen su origen en el grupo *Davenport*, un foro destinado a productores de libros electrónicos surgido a principios de la década de los 90. En 1993, se propuso la creación de una norma cuyo principal objetivo fuera posibilitar la fusión de índices impresos, que posteriormente evolucionó hacia otras estructuras (como tesauros y ontologías), hasta llegar a ser una herramienta considerada en el web para la organización, representación y gestión del conocimiento. Teóricamente, un Topic Map equivale a un índice creado para una colección de recursos disponibles en formato electrónico, cuyos elementos son:

- Una serie de tópicos o materias que se utilizan para describir el contenido de los recursos y facilitar su recuperación.
- Relaciones que establecen entre los tópicos.
- Una serie de recursos que se indizarán o describirán mediante los tópicos anteriores -ello equivale a decir que el recurso trata de o contiene información sobre ese tópico.

En tal sentido, un Topic Map sería similar a un catálogo de materias con los siguientes elementos:

- Un conjunto de encabezamientos que se pueden utilizar como descriptores.
- Relaciones explícita entre encabezamientos.
- Referencias a los libros a los que se han asignado las materias.

En la práctica, la implementación de los Topics Maps se diseñó para arquitectura SGML con notación de “*HyTime*” - la DTD del estándar ISO/IEC 13250:2000 perteneciente al diseño de Topics Maps se ha mantenido en HyTime-, no obstante, esta notación ha caído en desuso tras la aparición de XML. La Norma ISO /IEC 10744:1992 *Information technology - Hypermedia Time-Based Structuring Language* (Hytime) define el lenguaje y modelo subyacente para la representación de hiperdocumentos que enlazan y sincronizan información estática y dinámica (dependiente del tiempo) contenida en documentos convencionales o multimedia y otros objetos de información. En este caso, las causas del desplazamiento de la notación de HyTime se deben a la creación de una DTD para crear Topic Maps en XML, denominada XTM (*XML Topic Map*) y al auge que ha obtenido este. En cualquier caso, actualmente pueden encontrarse ejemplos de ambos lenguajes en la web.

Un *topic* es el elemento principal de un Topic Map, es el término que expresa determinado concepto o idea (o *subject*, como se denomina en el estándar). Ejemplos de *topics* pueden ser "Europa", "persona", "idea" o "continente". Básicamente, los *topics* tienen tres características principales: su denominación (*names*), sus apariciones u ocurrencias (*occurrences*) y su función en las asociaciones (*role associations*). Un topic puede tener varias denominaciones, pero debe estar representado por una forma base (*base name*). La denominación es un elemento obligatorio y representa la forma normal de hacer mención al *topic*. Las apariciones son enlaces a recursos informativos, como puede ser una página web.

Son elementos opcionales que son relevantes a un *topic* determinado, es decir, un *topic* puede tener cero o muchas ocurrencias.

Los recursos informativos pueden ser de muchos tipos: una cita de un texto, una definición, una ley, un artículo, etcétera. Cada uno de los diferentes tipos documentales puede agruparse mediante *roles de ocurrencia* (diccionario, página web, imagen). Es importante que la mayoría de estas ocurrencias suelen ser externas al Topic Map, una situación análoga a la que siempre ha existido entre un tesoro y la indización documental realizada con ese tesoro.

La asociación es un enlace que establece una relación entre dos o más *topics*. Una forma de verlo es tomar una frase y considerar los sustantivos de esas frases como *topics* y emplear los términos de unión entre los *topics* para denominar la asociación como por ejemplo: "las piñas *están formadas por* piñones"

Ahora bien, la especificación Topic Map se completa con un mecanismo de seriado, que se refiere a cómo codificarlo y escribirlo por medio de un formato legible y procesable por máquina. En este sentido, se han propuesto dos alternativas. Inicialmente, la norma ISO 13250 definió una DTD en SGML para su codificación, intercambio y procesamiento. Posteriormente, la generalización de XML hizo que se adaptase la DTD inicial a este lenguaje.

El resultado de este trabajo fue la especificación XML Topic Map, publicada en febrero de 2001 y, a su vez, la norma ISO 13250 fue revisada para incluir, como parte del estándar, la versión XML de la DTD. Ambos proyectos se centran en la definición conceptual de los Topics Maps y en su representación con lenguajes de marcas. "Actualmente, se encuentran en fase de desarrollo algunos estándares para la consulta y representación de los Topic Maps":

- La especificación TMQL (*Topic Map Query Language*): tiene como propósito definir un lenguaje para interrogar el contenido de los Topics Maps.

- La especificación TMCL (*Topic Map Constraint Language*): permitirá definir restricciones sobre el contenido expresado en el Topic Map.

El estado actual de estas especificaciones se puede consultar en: <www.isotopicmaps.org>

"Los Topic Maps constituyen una de las principales propuestas para la visualización de la web semántica". 37 Existe una problemática obvia en este tema cuando se tiene en cuenta que un Topic Map puede tener cientos de miles de asociaciones de diferentes tipologías. Actualmente, entre las representaciones propuestas destacan los *árboles*, los *browsers o navegadores* y los *gráficos*. A la hora de analizar estas representaciones, se debe considerar que muestren, tanto información local de los *topics* que interesan al usuario, como información sobre la localización de esos *topics* en su conjunto. En el caso de los gráficos, por ejemplo, se muestra un nodo principal y los nodos más próximos (fig. 6).



Fig. 6. Representación gráfica del metabuscador Kartoo

La figura muestra un metabuscador francés, salido a la luz en el año 2001 que, tras recopilar los resultados de 14 buscadores, elimina los duplicados, clasifica los documentos web según su pertinencia con respecto a la consulta y muestra la información final en una interfaz cartográfica. Kartoo presenta los resultados en forma de mapa inteligente, similar a los sistemas de representación neuronal, en los cuales una vez efectuada la consulta se pueden visualizar los resultados en forma de una lista de documentos. Dentro del mapa, cada sitio se representa con una esfera enlazada a otras de contenido relacionado y el tamaño de cada esfera es significativo de su relevancia, así como su aproximación al centro del mapa. Entre las esferas se incluyen enlaces que muestran que su contenido se encuentra relacionado por tratar aspectos comunes, es decir, cada enlace (tópico) muestra un concepto que indica el tema compartido por las esferas vinculadas.

Por otra parte, los Topic Maps se proponen para modelar redes semánticas. Por ejemplo, si se establece una relación entre un Topic Map y un tesoro se puede apreciar que los tesauros no son más que un caso particularmente simple del primero, donde tan solo existen tres tipos de relaciones (jerarquía, sinonimia

y relacionado).

Al definir el modelado inicial del Topic Map y un tesoro, las semejanzas no son tan evidentes. Una de las diferencias con los tesauros está más en el momento histórico en que cada propuesta tuvo lugar, así los tesauros son normalmente productos incorporados a una única organización de forma centralizada y cuya presentación en formato electrónico es sólo una característica extra, pero no un requisito. Los Topics Maps son productos descentralizados que mejoran con la cooperación de distintas organizaciones y que tienen esencialmente un formato electrónico.

Asimismo, los Topic Maps tienen una estrecha conexión con RDF porque, según Le Grand y otros autores, ambos están destinados a ser complementarios. La diferencia es que un Topic Map es independiente de que existan recursos de información o no, y permite en ocasiones una navegación entre conceptos. Por otra parte RDF, según *Sigel*, está más centrado en la descripción de recursos de información y no trata la semántica de las relaciones entre los meta-datos. Por lo tanto, las ventajas que tienen los Topics Maps para la representación del conocimiento en la web semántica, se resumen de la siguiente manera:

- Los Topic Maps pueden dar una semántica a elementos que están en el web, al organizarlos y describirlos, pero sin modificarlos.
- Presencia de perfiles de usuarios, p ermiten adaptarse a distintas comunidades para compartir recursos de información.
- Navegabilidad e inferencia mediante estructuras semánticas: Mejora no sólo la recuperación de información, sino también la gestión del conocimiento y el mantenimiento de los Topic Maps. En este punto, es también interesante recordar la independencia de los recursos de información que tiene un Topic Map, al poder navegar por nodos que no tengan ningún recurso asociado
- Fusión con otras estructuras de conocimiento, que permite una gestión descentralizada.
- Buena escalabilidad y compatibilidad para adaptarse al creciente número de recursos de información.

Entre las posibles desventajas, residen más en la poca madurez del estándar que en otros aspectos. Así se han revisado distintos aspectos como:

- La necesidad de mejorar la capacidad de integrar las propiedades de las dentro del esquema de definición de *Topic Maps* .
- La línea de investigación emergente sobre la forma de asegurar la consistencia de los *Topic Maps* .
- Las restricciones necesarias relativas a su validación y corrección.

El esfuerzo que supone la creación de un Topic Map realmente se compensa si se piensa en las posibilidades que ofrecen en cuanto a capacidad de indización a texto completo delegados a aplicaciones informáticas de carácter distribuido. Si bien estas aplicaciones son poco conocidas en nuestro ámbito académico, constituyen buenos ejemplos del potencial que ofrecen para la organización y la recuperación de información. Algunas de ellas, como *Ontopia* (<http://www.ontopia.net>), *Empolis* (<http://www.empolis.com>), *Mondeca* (<http://www.mondeca.com>) e *Infoloom* (<http://www.infoloom.com>)”, comparten características y funcionalidades; en todos los casos los Topics Maps se utilizan como una interfaz para indizar, interrogar y navegar repositorios de información.

AGENTES INTELIGENTES

Los agentes de software, denominados en el ámbito anglosajón *softbots o killer apps* , constituyen una de las áreas clave en el desarrollo de la inteligencia artificial, aunque intervienen otras áreas del conocimiento como la sociología, La informática, la lógica, la lingüística y las telecomunicaciones, entre otras, ello ha llevado a proponer la creación de una nueva disciplina dentro de la Ingeniería de Software, la Ingeniería de Software Orientada a Agentes (*Agent-Based Software Engineering*), cuyo fundamento es solucionar el problema que supone la heterogeneidad de las situaciones en la que puede encontrarse un agente de software y solucionar este problema mediante la comunicación y la cooperación.

En la inteligencia artificial ha surgido un nuevo paradigma conocido como” paradigma de agente”, que ha tomado un gran auge entre los investigadores en el diseño y desarrollo de entidades que puedan actuar de forma autónoma y razonada. Los investigadores en el campo de los agentes computacionales han aportado varias definiciones al término, cada uno desde su perspectiva particular, fundamentada básicamente en la línea de investigación en la cual trabajan (inteligencia artificial, ingeniería de software, sistemas autónomos). No obstante, la aplicación de los agentes de software es del interés para especialistas en organización de información por las perspectivas que abren en lo que se refiere a la localización, identificación, relación, mantenimiento y selección de recursos de información.

Un agente de software puede definirse como “un programa autocontenido capaz de controlar su proceso de toma de decisiones y de actuar, basado en la percepción del ambiente, en la persecución de uno o varios objetivos”. En esencia, de la definición anterior es posible extraer algunas características que deben tener los agentes:

- Deben ser parte de un ambiente.
- Deben censar su entorno y actuar sobre él.
- Deben responder según los objetivos para los que se diseñaron.

Para ello, según *Jennings y Wooldridge*, son necesarios una serie de atributos o propiedades que lo definen como agente: 38

- **Autonomía:** Capacidad de actuar sin intervención directa de una persona u otro agente, y controlar sus propias acciones y estado interno.
- **Habilidad social:** Un agente debe ser comunicativo, incorporar habilidades para interactuar con otros agentes o incluso con alguna persona, para solicitar información o bien para exponer los resultados obtenidos de la ejecución de las tareas encargadas.
- **Reactividad:** Se refiere al hecho de que un agente debe poder censar el estado del ambiente dentro del cual se encuentra inmerso y, en función de esto, actuar para responder de manera adecuada a los cambios que produzcan.
- **Orientación por objetivos:** Un agente no sólo debe actuar ante los cambios detectados en el medio, sino que – además – debe trabajar en función de los objetivos para los cuales se diseñaron y las tareas que les fueron delegadas en cada momento.
- **Continuidad temporal:** Un agente es un proceso temporalmente continuo. A diferencia de un programa convencional del cual se conoce su inicio y fin, un agente debe ejecutarse hasta que se alcance el conjunto de objetivos solicitados, o bien, mientras su ciclo perdure y su usuario no desee detenerlo.

El ciclo de vida de un agente depende de sus características, de las tareas que realice y de los deseos de su usuario en cuanto al tiempo durante el cual debe ejecutarse. De ahí, la movilidad como capacidad del agente para viajar por toda la red, nodo a nodo, en busca de recursos que cumplan con su estrategia de búsqueda. La movilidad no es una propiedad indispensable para un agente, sino que modifica la forma en que cumple con sus objetivos; en este caso, puede recurrir a los recursos que ofrece una red de computadoras, y aportar una nueva forma de computación distribuida. Según esto, *Berney* propone una clasificación de agentes de acuerdo con sus líneas de investigación y desarrollo: 39

- **Agentes de interfaz:** Un agente de interfaz es un software quasi-inteligente que asiste a un usuario cuando interactúa con una o más aplicaciones. La motivación es que se les pueda delegar tareas complejas y laboriosas. Son asistentes personales que reducen el trabajo ante la sobrecarga de información, por ejemplo, el filtrado de los mensajes de correo electrónico o la recuperación de archivos en Internet.
- **Agentes de colaboración:** Constituyen sistemas multiagentes, es decir, existe más de un agente dedicado a satisfacer los requerimientos de sus usuarios. Para ello, es necesario contar con esquemas de comunicación entre agentes que posibiliten la cooperación y el intercambio de conocimiento.
- **Agentes móviles:** Los agentes móviles son aplicaciones capaces de viajar por una red de computadoras, interactuando con servidores externos y recolectar información en nombre de su dueño y retornar luego

de completar las tareas establecidas previamente.

- Agentes de recuperación de información: El objetivo principal de los agentes dedicados específicamente a la recuperación de información es obtener información para el usuario. La motivación para su construcción es que con el crecimiento vertiginoso de Internet, la cantidad de información accesible supera el tiempo de disponible para analizarla.

“El papel del agente inteligente en el proceso de recuperación semántica de información no debe confundirse con la de un buscador inteligente”. 40 Un buscador inteligente se aprovechará del enriquecimiento semántico de los recursos web para mejorar, principalmente, en la precisión y la recuperación de información. , aunque su funcionamiento se basará, como los actuales buscadores, en la previa indización de todos aquellos recursos susceptibles de recuperarse. Por otra parte, el agente inteligente recorrerá el web por medio de enlaces entre recursos (taxonomías, tesauros, ontologías) en busca de aquella información que le sea solicitada y puede además interactuar con el entorno para el cumplimiento de tareas encomendadas mediante la utilización de esquemas de metadatos codificados en esquemas RDF (*RDF Schema*) que pudiesen procesar y reutilizar ontologías.

La web semántica proyecta como uno de sus fundamentos el desarrollo de sistemas de información basados en el conocimiento de agentes inteligentes para lograr la interoperabilidad, no sólo sintáctica y semántica, sino entre aplicaciones independientemente de la acción de los usuarios de la red. Sustentados sobre la base de la heterogeneidad, constituyen nuevos métodos para la organización y la recuperación de información en entornos distribuidos.

WEB SERVICES

El costo de desarrollar software siempre es muy alto y la diversidad de las plataformas es una realidad desde el inicio de la Informática. De hecho, conforme más complejas fueron las aplicaciones que las empresas demandaban, más caras era desarrollarlas. Para enfrentarse a esta situación, se han seguido diferentes líneas de implementación, todas ellas dirigidas a reutilizar las aplicaciones desarrolladas.

Una de estas propuestas fue y es la estandarización de lenguajes de programación de forma que si cualquier desarrollador escribía un programa en “C”, sólo necesitaría un compilador de C en la plataforma específica en la que quisiera ejecutar la aplicación. Esto, en realidad, es difícil de hacer funcionar, porque surgieron pequeñas diferencias y extensiones de C que hacían difícil transportar una aplicación entre diferentes plataformas. C, un lenguaje de programación desarrollado en 1972 por el estadounidense *Dennis Ritchie* en los Laboratorios Bell, considerado por muchos un lenguaje ensamblador más independiente de la máquina que un lenguaje de alto nivel, su estrecha asociación con el sistema operativo UNIX, su enorme popularidad y su homologación por el American National Standards Institute (ANSI) lo han convertido tal vez en lo más cercano a un lenguaje de programación estandarizado en el sector de las microcomputadoras y estaciones de trabajo.

Otra posibilidad fue la que ha desarrollado Sun Microsystems con “Java”. Se programa para una

plataforma, pero para una plataforma virtual, en este caso, para la máquina virtual Java (*Java Virtual Machine*), y en cada plataforma real se implementa una máquina virtual de Java, que será la encargada de ejecutar las aplicaciones escritas en Java.

Java es lenguaje de programación orientado a objetos desarrollado por la empresa Sun Microsystems en 1995 y que se ha extendido ampliamente en el web. Es un lenguaje de alto nivel y propósito general similar a C++, con marcadas características de seguridad y transportabilidad. Este lenguaje define una máquina virtual independiente de la plataforma donde se ejecuta, que procesa programas, llamados Applets, descargados desde el servidor web. Además, debido al modo de ejecución de los Applets, este lenguaje es muy seguro frente a la presencia y ataque de virus informáticos.

Las implementaciones son específicas de cada plataforma, pero al ser todas las máquinas virtuales exactamente la misma, los programas escritos en Java deben poder ejecutarse sin ningún problema en todas las plataformas que tengan una máquina virtual de Java.

Esta apuesta ha demostrado ser muy útil y funcionar bastante bien. Sin embargo: ¿qué pasa si se tienen varias aplicaciones desarrolladas en lenguajes propietarios o en plataformas específicas y que interactúen entre ellas? El costo de elegir un único lenguaje o plataforma y migrarlo al mismo es descabellado en la mayoría de las situaciones, y es aquí donde los Web Services, así como otras tecnologías, pueden ser de gran utilidad.

Con los Web Services se pueden reutilizar desarrollos empleados sin importar la plataforma en la que funcionan o el lenguaje en el que están escritos. Los Web Services se constituyen en una capa adicional a estas aplicaciones de forma que pueden interactuar entre ellas, a partir del uso de tecnologías estándares para comunicarse, desarrolladas en el contexto de Internet por organismos importantes de estandarización como: ANSI, IEEE (*Institute of Electric and Electronics Engineers* .), ISO y W3C.

Actualmente, gran parte de los presupuestos destinados al diseño y mantenimiento de sistemas de información se dedican a resolver problemas de integración entre aplicaciones, de forma que puedan comunicarse independientemente de su plataforma operativa en tiempo real. Muchas de las compañías líderes del mercado de software han concebido un nuevo método mediante el cual solucionar el problema de la heterogeneidad e interoperabilidad entre aplicaciones web mediante el diseño y mantenimiento de los Web Services. "*Un Web Service es un componente de software que se comunica con otras aplicaciones mediante la codificación de los mensaje en XML y el envío de estos mensaje con protocolos estándares de Internet como el Hyper -Text Transfer Protocol (HTTP)*". 41

El desarrollo de servicios web o *Web Services* como aplicaciones en entornos distribuidos y descentralizados, básicamente utiliza las siguientes tecnologías: 41

- Un formato que describa la interfaz del componente (sus métodos y atributos) basado en XML. Por lo general, este formato es el WSDL (*Web Service Description Language*).

- Un protocolo de aplicación basado en mensajes y que permita que una aplicación interactúe (use, llame, ejecute) con el Web Service. Por lo general, este protocolo es SOAP (*Simple Object Access Protocol*).
- Estructuras parecidas a las bases de datos que organizan los servicios web; así como especificaciones que definen un camino para publicar y descubrir información sobre los servicios Web mediante el estándar UDDI (*Universal Description, Discovery and Integration*).
- Un protocolo de transporte que se encargue de transportar los mensajes por Internet. Por lo general, este protocolo de transporte es HTTP (*Hyper-Text Transfer Protocol*) que es exactamente el mismo que se usa para navegar por el web.

Los Web Services no son, por tanto, aplicaciones con una interfaz gráfica con las que las personas puedan interactuar, sino que son software accesibles en Internet o en redes privadas que utilizan tecnologías de Internet en lugar de otras aplicaciones. De esta forma, pueden desarrollarse aplicaciones que hagan uso de otras, que estén disponibles en Internet e interactúen con ellas.

En este sentido, un aspecto de suma importancia para la puesta en marcha y mejora de la web semántica es la existencia de herramientas de soporte y la re-utilización de ontologías que puedan utilizar los Web Services para recuperarlas, mediante una plataforma que entornos de desarrollo de aplicaciones para la web semántica y un módulo de conexión a servicios web para la búsqueda y recuperación de ontologías.

Definido sobre esta base, empresas como Hewlet Packard y la National University of Ireland trabajan desde el año 2002 en una plataforma orientada al descubrimiento, descripción y mediación entre Web Services en el marco de la web semántica conocida como *Semantic Web Enabled Web Services* . Además, el *Computer Science and Electrical Engineering Department* de la Universidad de Maryland, desarrolla un proyecto de investigación iniciado en el año 2004 con un período de duración de dos años, que consiste en un buscador temático llamado Swoogle, con una interfaz de consulta muy sencilla que recupera ontologías codificadas en RDF, OWL y DAML+OIL mediante esta plataforma de intercambio de Web Services. (fig. 7)

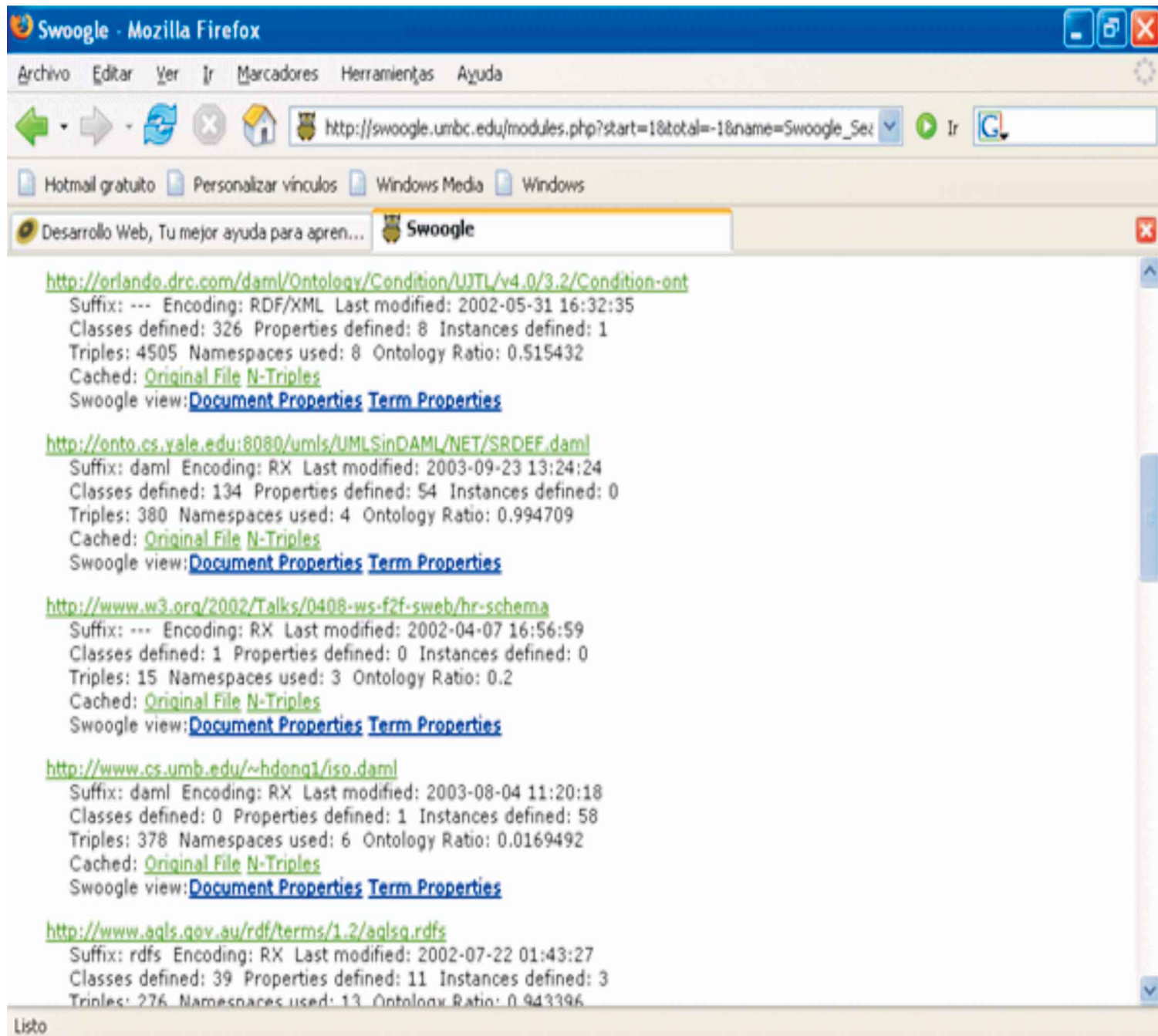


Fig. 7. Buscador implementado para recuperar ontologías en la web semántica

Por otra parte, es válido aclarar que los Web Services utilizan las mismas tecnologías con las que se conciben muchas de las aplicaciones web actuales como: *Hyper-text Pre-Processor* (php), *Active Server Page* (asp), *Java Server Page* (jsp), entre otras, por lo que pueden tener los mismos problemas de seguridad que estas tecnología enfrentan. A ello hay que añadir que actualmente los Web Services no son capaces de ofrecer todo el conjunto de características avanzadas de muchos sistemas de integración de aplicaciones, por ejemplo: gestión de la capa de comunicación, coordinación de transacciones, mapeo de datos, gestión de procesos de negocios y seguridad, aunque son cuestiones sobre las que se trabaja.

FIRMAS DIGITALES Y REDES DE CONFIANZA

Uno de las componentes fundamentales de la web semántica serán las firmas digitales. Estas se definen como “bloques de datos cifrados que las computadoras y agentes de software podrán utilizar para verificar que la información adjunta fue proporcionada por una fuente fiable”. 42 Al aplicar un algoritmo de cifrado sobre un documento o mensaje a firmar, se obtiene un bloque de datos que representa un resumen del documento o mensaje. La firma digital se obtiene mediante un *Sistema de criptografía de clave pública*, que se basa en el empleo de dos tipos de claves: pública y privada, todo lo que se cifre mediante una clave pública solo puede ser descifrado a por medio de su correspondiente clave privada y viceversa.

Las ventajas que confiere la firma digital a los recursos, documentos o mensajes del web, son principalmente:

- Identificación: Posibilidad de determinar la identidad del emisor o autor del recurso.
- Integridad: Facilidad para detectar la manipulación o alteración ilícita del recurso.

La privacidad de los datos personales que se manejan en Internet es una preocupación constante para empresas, gobiernos, medios de comunicación y el público en general. Para solucionar este problema, surgieron grupos de trabajo e iniciativas dirigidas a estandarizar estos problemas como los grupos de trabajo XML Encryption y XML Signature, este último fruto del esfuerzo conjunto entre el W3C y el IETF (Internet Engineering Task Force), quienes han desarrollado una especificación, con sintaxis XML, para la representación de firmas digitales en recursos web que alcanzó el grado de Recomendación del W3C en febrero de 2002.

XML Signature asegura la integridad de partes de documentos XML transportados, así como proporciona la autenticación de mensajes y servicios de autenticación de firma para datos de cualquier tipo, tanto si se encuentra en el XML que incluye la firma o en cualquier otra parte. Puede aplicarse a cualquier contenido digital (objeto de datos), incluido XML por medio de la asociación de claves con los datos de consulta. XML Signature representa un sistema que mediante una firma digital permite ofrecer autenticidad de los datos. Con la firma digital se confirma la identidad del emisor, la autenticidad del mensaje e integridad.

En el caso de XML Encryption, es un lenguaje cuya función principal es asegurar la confidencialidad de partes de documentos XML por medio de la encriptación parcial del documento transportado. XML Encryption puede aplicarse a cualquier recurso web, incluido contenido que no es XML.

Otro de los esfuerzos orientados a la seguridad es la conocida como P3P o Plataforma de Preferencias de Privacidad (*Platform for Privacy Preferences*), que nace ante la necesidad de garantizar la privacidad en una web cada vez más extensa, que alcanzó el grado de Recomendación por el W3C en abril de 2002. P3P es un lenguaje estándar que ofrece a los usuarios una forma sencilla y automatizada de controlar en

mayor medida el uso que se hace de su información personal en los sitios web que visitan. A su vez posibilita:

Desarrollar herramientas y servicios que ofrezcan a los usuarios un mayor control sobre la información personal que se maneja en Internet y, al mismo tiempo, aumentar la confianza entre los servicios web y los usuarios.

Mejorar el control del usuario al colocar políticas de privacidad donde los usuarios pueden encontrarlas, en un formato en el que los usuarios pueden entender y, lo más importante, con la posibilidad de que el usuario actúe sobre lo que ve.

A los usuarios web, facilidad y regularidad a la hora de decidir si quieren o no, y bajo qué circunstancia, revelar información personal.

Permite a los sitios web trasladar sus prácticas de privacidad a un formato estandarizado y procesable por dispositivos -basado en XML - que puede ser recuperado de forma automática y que además puede interpretarse fácilmente por los navegadores web. Sin embargo, “aunque las firmas digitales resultan de utilidad para describir la autenticidad de las relaciones de autoría entre individuos y recursos, no describen la confianza de los usuarios y agentes inteligentes sobre dichos autores y, por lo tanto, sobre el contenido de sus recursos”. 43

La solución puede estar en el concepto de una *Web de Confianza* que basa su funcionamiento sobre la base de la confianza o fiabilidad entre usuarios y agentes de software en la red, definido como “el grado en que un agente considera un aserto como verdadero para un contexto dado”. 44 Cada usuario, identificado por un URI, expresaría su grado de confianza o desconfianza sobre otros individuos, que a su vez harían lo mismo sobre otros, lo que daría como resultado extensas e interoperables redes de confianza procesables por agentes inteligentes.

Para la web semántica se han puesto en práctica aplicaciones que propician el análisis de redes sociales que toman como premisa un vocabulario RDF para describir información sobre los individuos y sus relaciones en una red social, conocido como FOAF (*Friend OF A Friend*), que se extiende con el objetivo de expresar el nivel de confianza entre los individuos de la red para finalmente mostrar su implementación práctica en áreas como el correo electrónico y la mensajería instantánea.

El uso conjunto de firmas digitales y redes de confianza, aplicadas a las reglas de inferencia con la cuales operan los agentes de software, pueden solucionar los problemas de autenticación y validación de permisos de usuarios, así como entre estos y los computadoras. El problema de la centralización de los mecanismos de seguridad puede resolverse mediante el uso generalizado de redes de confianza y bloques de datos cifrados, que harían innecesario el empleo de criptografía en la mayoría de los casos y, consecuentemente, de las Autoridades de Certificación, en un entorno caracterizado por la descentralización en el que pretende establecerse la web semántica.

MODELO CONSTRUCTIVO DE LA WEB SEMÁNTICA

Una vez analizadas todas las tecnologías y lenguajes necesarios para la implementación de la web semántica, se puede esquematizar de forma gráfica cómo esta infraestructura, dividida en varias capas o niveles, se complementa para lograr dotar al web actual de semántica desde el punto de vista teórico-funcional (fig. 8). Este diagrama, presentado por *Berners-Lee* en la *XML Conference* del año 2000, puede servir como aproximación visual al conjunto de tecnologías que forman el esquema de capas mencionado, cuyos componentes son:

- Estándares para la localización de recursos de información en el web de forma inequívoca y única como son los URIs (*Uniform Resource Identifiers*) y la norma internacional Unicode para la codificación de caracteres a nivel internacional.
- XML (*eXtensible Markup Language*), como base sintáctica para la estructuración del contenido en el web, así como el empleo de espacios de nombres (*Namespaces*) para asociar con precisión cada propiedad con el esquema que define dicha propiedad y esquemas (*XML Schema*) para definir qué elementos debe contener un documento XML, cómo están organizados, qué atributos y de qué tipo pueden tener sus elementos.
- Un modelo básico para establecer propiedades sobre los recursos, para el que se empleará RDF (*Resource Description Framework*), así como un modelo para definir relaciones entre los recursos por medio de clases y objetos, que se expresan mediante esquemas en RDF (*RDF Schema*).
- Lenguajes para la representación de ontologías que permitan la interoperabilidad y reutilización entre ontologías de diversos dominios del conocimiento en el web, cuya base se encuentra en *RDF Schema* .
- Una capa lógica que permita realizar consultas e inferir conocimiento, donde estarían las ontologías, agentes software y web services como estructuras para lograr interoperabilidad entre aplicaciones y sistemas de información heterogéneos.
- Una capa de seguridad que permita asignar niveles de fiabilidad a determinados recursos, de forma comprobable posteriormente por los agentes, para lo que se utilizarán firmas digitales y redes de confianza.

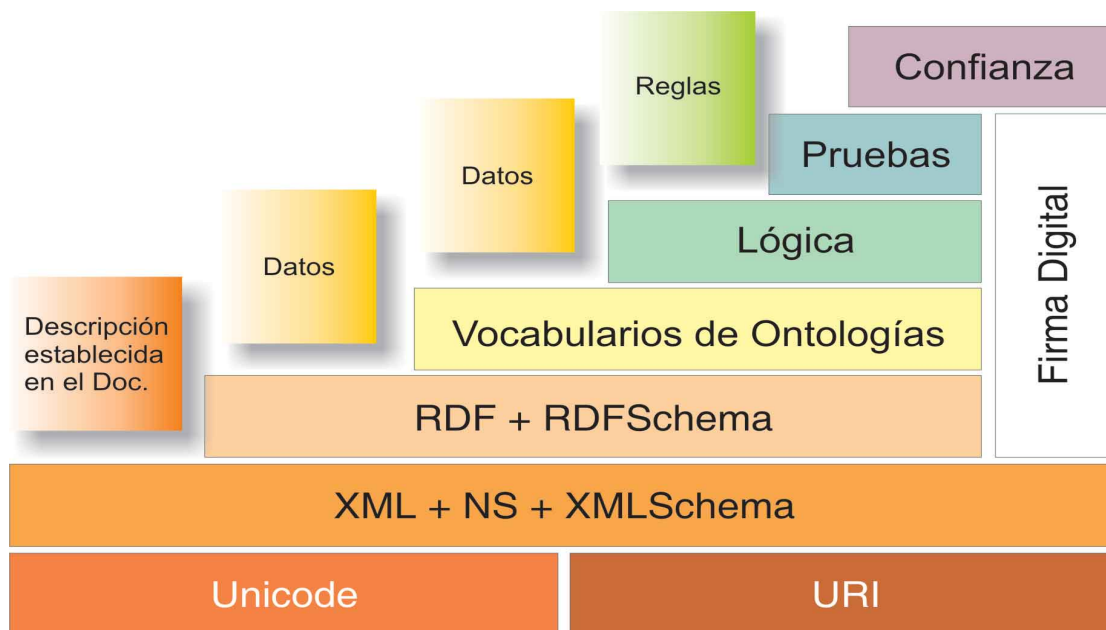


Fig. 8. Modelo de capas propuesto por Berners-Lee para la web semántica

Ossenbruggen, Hardman y Rutledge plantean que las primeras capas de la web semántica (URIs, XML, RDF), que tienen la función de describir, estructurar e identificar los recursos web, tienen soporte tecnológico en la web actual, mientras que las capas que pretenden dotar de lógica e inteligencia al web, no tienen, por el momento, soporte en el web actual (agentes y ontologías).⁴⁵ De igual manera, se considera que las capas necesarias para dotar de fiabilidad, confianza y seguridad al web se encuentran en discusión (firmas digitales y redes de confianza). Otro de los problemas que afronta este modelo es que cada capa mantiene su propia sintaxis y no todas extienden de la misma manera su semántica.

Analizada la propuesta de la web semántica, cabe cuestionarse qué mejoras supondría este modelo para el web actual. Mientras mayor sea el nivel de descripción semántica y estructuración lógica de los documentos en el web, mayor será la precisión obtenida en la recuperación de la información. De igual forma, la exhaustividad mejoraría gracias a que los axiomas definidos en las ontologías posibilitarán tanto a agentes como buscadores inteligentes, inferir nuevo conocimiento no explicitado previamente y recuperar un mayor número de recursos relevantes para determinada consulta.

CONCLUSIONES

El web ha cambiado profundamente la forma en la que nos comunicamos, hacemos negocios y realizamos nuestro trabajo. Se pueden realizar transacciones económicas por medio de Internet y tener acceso a millones de recursos, independientemente de nuestra ubicación geográfica e idioma. Todos estos factores han contribuido al éxito del web. Sin embargo, al mismo tiempo, estos factores que han propiciado su éxito, también han originado sus principales problemas: sobrecarga de información y heterogeneidad de fuentes de información con el consiguiente problema de interoperabilidad.

La web semántica ayudaría a resolver estos dos importantes problemas al permitir a los usuarios delegar

tareas específicas al software. Gracias a la semántica en el web, el software sería capaz de procesar su contenido, razonar con este, combinarlo y realizar deducciones lógicas para resolver problemas cotidianos automáticamente. En este sentido, después de realizar un análisis general con relación a los componentes y tecnologías que integran el modelo teórico-conceptual del proyecto de la web semántica, así como lograr un consenso entre la diversidad de criterios que en torno a esta iniciativa han tenido sus principales investigadores, puede afirmarse que:

La concepción de los procesos organizativos de la información como la clasificación y de la indización documental en ambientes automáticos y, específicamente en el web, tienden a integrar modelos estadísticos y lingüísticos (análisis sintáctico y semántico) en función del desarrollo de las técnicas de procesamiento del lenguaje natural, de la inteligencia artificial y en favor de la creación de sistemas de información mixtos que conjuguen el software con capacidad de razonamiento e inferencia, como paso previo para la mejora de las técnicas de recuperación de información.

La web semántica, como área interdisciplinar, proporcionará un salto cualitativo sobre el potencial del web, porque una de las principales ventajas de esta extensión del web actual radica en el desarrollo de aplicaciones con esquemas de datos comunes para el fomento de las transacciones entre empresas por comercio electrónico y búsqueda de información mediante inferencias.

El desarrollo de la web semántica proporcionará tres niveles fundamentales de interoperabilidad entre esquemas de datos y aplicaciones heterogéneos:

- Interoperabilidad sintáctica: mediante XML como metalenguaje de marcado para estructurar la información web.
- Interoperabilidad semántica: mediante RDF como modelo de datos genérico y contenedor de esquemas de datos específicos.
- Interoperabilidad estructural: mediante la unificación los contenidos semánticos por medio de ontologías que formalicen el conocimiento en forma consensuada y reutilizable.

Se necesita un lenguaje común basado en el web, con suficiente capacidad expresiva y de razonamiento para representar la semántica de las ontologías. Este constituye uno de los inconvenientes del actual modelo de capas planteado por *Berners Lee*, porque no todos sus niveles extienden la semántica con el mismo grado de interoperabilidad.

En este sentido, los futuros agentes de software se diseñarán para explotar el conocimiento del web contenido no sólo en las ontologías, sino en otras estructuras de organización de la información como pueden ser las taxonomías, tesauros y los Topic Maps.

REFERENCIAS BIBLIOGRÁFICAS

1. Maller P, Dahaes V. Web Semántica: el salto evolutivo de la Web. Disponible en: <http://www.proyectoweb.org/boletines/034-marz03.html> [Consultado: 20 de septiembre del 2004].
2. Castells P. Aplicación de técnicas de la web semántica. Disponible en: <http://giig.ugr.es/~mgea/coline02/Articulos/pcastells.pdf> [Consultado: 2 de febrero del 2005].
3. Matthews B, Wilson M, Brickley D. Semantic Web Advanced Development in Europe . ERCIM News 2002;(51):15-6.
4. Berners Lee T, Miller E. The Semantic Web lifts off, W3C. ERCIM News 2002;(51). Disponible: <http://www.w3.org/2002/03/key-free-trust> [Consultado: 23 de noviembre del 2004].
5. Berners Lee T, Handler J, Lassila O. The Semantic Web. Scientific American 2001;284(5).
6. Daconta MC, Obrst Leo J, Smith KT. The Semantic Web: A guide to the future of XML, Web Services and Knowledge Management. New York:: Wiley, 2003 p. 145.
7. Merriam- Webster Online. Disponible en: www.m-w.com [Consultado: 20 de marzo del del 2005].
8. Novak JD, Gowin DB. Learning How to Learn . New York: Cambridge University Press, 1984.
9. Roberts N. Historical studies in documentation: pre-history of information retrieval thesaurus. Journal of Documentation 1984;40(4):271-85.
10. Corcho O, Fernández López M, Gómez Pérez A. Methodologies, tools and languages for building ontologies: where is their meeting point? Data & Knowledge Engineering 2003;46(1):41-64.
11. García Jiménez A. Instrumentos de representación del conocimiento: tesauros versus ontologías. Anales de la Documentación 2004;(7):7.
12. Studer R, Benjamins R, Fensel D. Knowledge Engineering: Principles and Methods. Data and Knowledge Engineering 1998;25(1/2):161-97.
13. Neches R. Enabling technology for knowledge sharing. AI Magazine 1991;12(3):36-56.
14. Noy NF, McGuinness DL. Ontology Development 101: A Guide to Creating Your First Ontology. 2000. Disponible en : <http://www.cs.man.ac.uk/~carole/old/GGF%20Tutorial%20Stuff/ontology101.pdf> [Consultado: 11 de noviembre del 2004].

15. Guarino N. Formal ontology and information systems. En: Guarino N (ed.). *Formal Ontology in Informaton Systems* . Amsterdam: IOS Press, 1998.
16. Guerrero Bote V, Lozano Tello A. Vínculos entre las Ontologías y la Biblioteconomía y Documentación. En: *La Representación y la Organización del Conocimiento en sus distintas perspectivas: su influencia en la Recuperación de la Información*. Actas del IV Congreso ISKO-España EOCONSID'99, abril 22-24, Granada, España. Granada: ISKO-Facultad de Biblioteconomía y Documentación, 1999. p. 25-31.
17. Ding Y, Foo S. Ontology research and development. Part 1. A review of ontology generation. *Journal of Information Science* 2002;28(2):123-36.
18. Ding Y, Foo S. Ontology research and development. Part 2. A review of ontology mapping and evolving. *Journal of Information Science* 2002;28(5):375-88.
19. García Jiménez A. Instrumentos de representación del conocimiento: tesauros versus ontologías. *Anales de la Documentación* 2004;(7):7.
- Gómez Pérez A. (coord.). *A Survey on Ontology Tools*, OntoWeb deliverable 1.3.2002. Disponible en: http://www.aifb.uni-karlsruhe.de/WBS/ysu/publications/OntoWeb_Del_1-3.pdf [Consultado: 11 de noviembre del 2004].
20. Paling S. Converting a controlled vocabulary into an ontology. *Information Research* 2001;6(2). Disponible en: <http://informationr.net/ir/6-2/paper94.html> [Consultado: 5 de noviembre del 2004].
21. Triángulo semántico o semiótico. Disponible en: <http://culturitalia.uibk.ac.at/hispanoteca/Lexikon%20der%20Linguistik/sa/SEMIOTISCHES%20DREIECK%20%20%20Tri%C3%A1ngulo%20sem%C3%A1ntico%20o%20semi%C3%B3tico.htm> [Consultado: 5 de noviembre del 2004].
22. Daconta MC, Obrst LJ, Smith KT. *The Semantic Web: A guide to the future of XML*. Web Services and Knowledge Management New York: Wiley, 2003.
23. Minsky ML. *Semantic Information Processing* Cambridge: MIT Press, 1968.
24. Branchman RJ. *Structural Paradigm for Representing Knowledge*. Harvard: Harvard University, 1977.
25. ISO. ISO 8879. *Information processing -- Text and office systems -- Standard Generalized Markup Language (SGML)*. Geneva: ISO, 1986.

26. University of Chicago. The Chicago Manual of Style. 15 th ed. Chicago: University of Chicago Press, 2005.
27. Van Herwijnen E. Practical SGML. 2 ed. Boston: Kluwer Academic Publishers, 1994.
28. Irons Walch V. Standard for Archival Descriptions: A Hanbook. Disponible en: <http://www.archivists.org/catalog/stds99/chapter3.html> [Consultado: 20 de marzo del 2005].
29. Peis E, Ruíz Rodríguez AA. EAD (Encoded Archival Description): Desarrollo, estructura, uso y aplicaciones. Disponible en: <http://www.hipertext.net/web/pag223.htm> [Consultado: 10 de enero del 2005].
30. Burnage G, Dunlop D. Encoding the British National Corpus. Disponible en: <http://www.natcorp.ox.ac.uk/using/papers/Burnage93a.htm> [Consultado: 15 de enero del 2005].
31. Burnard L, Light R. Three SGML metadata formats: TEI, EAD, and CIMI. A Study for BIBLINK Work Package 1.1. 1996. Disponible en: <http://hosted.ukoln.ac.uk/biblink/wp1/sgml/> [Consultado: 15 de enero del 2005].
32. Rusty Harold E, Means W. XML. Madrid: Anaya Multimedia, 2005.
33. Fernández M, Simeón J, Walder P. XML Query Language: Experiences and Examples. Disponible en: <http://www.w3.org/1999/09/ql/docs/xquery.html> [Consultado: 9 de marzo del 2005].
34. Díaz O, Iturroz J, Ibáñez F. Integración, navegación y presentación: experiencias utilizando XML. Novótica 2000;146(4):12-9.
35. Miller E. An Introduction to the Resource Description Framework. D-Lib Magazine 1998;(5). Disponible en: <http://www.dlib.org/dlib/may98/miller/05miller.html> [Consultado: 15 de abril del 2005].
36. Senso JA. Herramientas para trabajar con RDF. El profesional de la Información 2003;12(2):132-8.
37. Le Grand B and Soto M. Visualisation of the Semantic Web: Topic Maps Visualisation. En: IEEE Computer Society. Proceedings of Six International Conference on Information Visualisation, July 10-12, London, England. London: IEEE Computer Society, 2002.
38. Jennings N, Wooldridge M. Agent Technology - Foundations, Applications, and Markets, New York: Springer-UNICOM, 1998.
39. Berney B. Software Agents. A Review. Manchester: Manchester Metropolitan University, 1996.

40. Hassan Montero, Y. Ontologías, taxonomías y agentes: recuperación “semántica” de la información. En: Universidad Carlos III. Actas de las II Jornadas de Tratamiento y Recuperación de Información (JOTRI - 2003), septiembre 8-9, Madrid, España. Madrid: Universidad Carlos III, 2003.
41. Fensel D, Bussler C. Semantic Enabled Web Services. 2002. Disponible en: <http://www.swsi.org/resources/wsmf.pdf> [Consultado: 20 de marzo del 2005].
42. Hendler J, Golbeck J, Parsia B. Trust Networks on the Semantic Web. 2003. Disponible en: <http://www.mindswap.org/papers/CIA03.pdf> [Consultado: 15 de marzo del 2005].
43. Reagle JM. Key Free Trust in the Semantic Web: Finding Bacon's Key.
Disponible en: <http://www.w3.org/2002/03/key-free-trust> [Consultado: 23 de abril del 2005].
44. Hendler J, Golbeck J, Parsia B. Trust Networks on the Semantic Web. En: Proceedings of Cooperative Intelligent Agents, August 27-29, Helsinki, Finland. Helsinki: s.e, 2003.
45. Van Ossenbruggen J, Hardman L, Rutledge L. Hypermedia and the Semantic Web: A Research Agenda. [Journal of Digital Information](#) 2002;3(1). Disponible en: <http://jodi.ecs.soton.ac.uk/Articles/v03/i01/VanOssenbruggen/> [Consultado: 2 de mayo del 2005].

Recibido: 25 de noviembre del 2005.

Aprobado: 6 de diciembre del 2005.

[1-Licenciado en Bibliotecología y Ciencia de la Información. Red Telemática de Salud en Cuba \(Infomed\). Centro Nacional de Información de Ciencias Médicas-Infomed.](#)

[2-Licenciado en Bibliotecología y Ciencias de la Información. Facultad de Comunicación. Universidad de La Habana.](#)

Lic. Keilyn Rodríguez Perojo. Red Telemática de Salud en Cuba. Centro Nacional de Información de Ciencias Médicas-Infomed. Calle 27 No. 110 e/ N y M, El Vedado. Plaza de la Revolución. Ciudad de La Habana. Cuba. Correo electrónico: keylin@infomed.sld.cu

Ficha de procesamiento

Clasificación: Artículo monográfico.

Términos sugeridos para la indización

Según DeCS 1

INTERNET; ALMACENAMIENTO Y RECUPERACIÓN DE LA INFORMACIÓN .

INTERNET; INFORMATION STORAGE AND RETRIEVAL .

Según DeCI 2

INTERNET; WWW; CLASIFICACIÓN; INDIZACIÓN, RECUPERACIÓN DE LA INFORMACIÓN ;
TESAUROS; METADATOS; TECNOLOGÍA DE LA INFORMACIÓN , TECNOLOGÍA DE LA
COMUNICACIÓN.

INTERNET; WWW; CLASSIFICATION; INDEXING; INFORMATION RETRIEVAL; THESAURY;
METADATA; INFORMATION TECHNOLOGY; COMMUNICATION TECHNOLOGY.

1 BIREME. Descriptores en Ciencias de la Salud (DeCS). Sao Paulo: BIREME, 2004.

Disponible en: <http://decs.bvs.br/E/homepagee.htm>

2 Díaz del Campo S. Propuesta de términos para la indización en Ciencias de la Información.
Descriptores en Ciencias de la Información (DeCI). Disponible en: <http://cis.sld.cu/E/tesauro.pdf>