

Universidad de Granada

Facultad de Comunicación y Documentación

Departamento de Biblioteconomía y Documentación



**MODELO DE SERVICIO SEMÁNTICO-DIFUSO  
DE DIFUSIÓN SELECTIVA DE INFORMACIÓN (DSI)  
PARA BIBLIOTECAS DIGITALES**

MEMORIA DE TESIS PRESENTADA POR

**JOSÉ MANUEL MORALES-DEL-CASTILLO**

COMO REQUISITO PARA  
OPTAR AL GRADO DE DOCTOR  
EN DOCUMENTACIÓN

Granada, junio de 2008

Editor: Editorial de la Universidad de Granada  
Autor: José Manuel Morales del Castillo  
D.L.: GR.1928-2008  
ISBN: 978-84-691-5801-2



Universidad de Granada

Facultad de Comunicación y Documentación

Departamento de Biblioteconomía y Documentación



**MODELO DE SERVICIO SEMÁNTICO-DIFUSO  
DE DIFUSIÓN SELECTIVA DE INFORMACIÓN (DSI)  
PARA BIBLIOTECAS DIGITALES**

MEMORIA DE TESIS PRESENTADA POR

**JOSÉ MANUEL MORALES-DEL-CASTILLO**

PARA OPTAR AL TÍTULO DE DOCTOR EN DOCUMENTACIÓN

DIRECTORES:

**DR. EDUARDO PEIS REDONDO**

**DR. ENRIQUE HERRERA-VIEDMA**

Fdo. J.M. Morales del Castillo

Fdo. Eduardo Peis

Fdo. Enrique Herrera

Granada, junio de 2008



---

## AGRADECIMIENTOS

*Todos nos necesitamos los unos a los otros. La gente con la que convivimos y que forma parte de nuestra familia, de nuestro círculo afectivo o profesional de alguna manera siempre es partícipe de lo que conseguimos en la vida.*

*Por esta razón, pecaría de ingrato si no hiciera un reconocimiento público a todas aquellas personas que han contribuido a que esta tesis sea una realidad. A todas ellas les quiero dar las gracias.*

*Gracias a Eduardo y Enrique, mis directores de tesis, por su apoyo incondicional y por haber creído en mí y en mi trabajo.*

*Gracias a mi padre y a mi madre, por sus desvelos y sacrificios para conseguir darme una formación y unas oportunidades de las que ellos se vieron privados.*

*Y gracias también a Ana Belén y Loren por tenderme su mano en el momento que más lo necesitaba; a Migue por animarme a continuar en la brecha; a compañeros, como Antonio, por sus sabios consejos; a Yusef por las charlas de cafetería que hemos compartido; y gracias, en definitiva, a todos aquellos que me han apoyado durante todo este tiempo.*

*Todos vosotros formáis parte de este trabajo.*



---

# ÍNDICE GENERAL

## Capítulo 1: Introducción

1.1 Justificación.....	1
1.2 Objetivos.....	10
1.3 Metodología.....	11
1.3.1 Investigación de campo.....	11
1.3.2 Investigación documental.....	11
1.3.2.1 Análisis de cocitación de autores.....	14
1.4 Estructura del trabajo.....	22

## Capítulo 2: Web Semántica y bibliotecas digitales

2.1 Introducción.....	25
2.2 ¿Qué es la Web Semántica?.....	32
2.2.1 Capa sintáctica.....	36
2.2.2 Capa semántica.....	39
2.2.3 Capa ontológica.....	44
2.2.4 Capa lógica.....	46
2.2.5 Capas de prueba y confianza .....	49
2.2.6 Toda la maquinaria en funcionamiento.....	51
2.3 Grupos de trabajo y proyectos sobre Web Semántica.....	53
2.3.1 Grupos y proyectos del W3C.....	55
2.3.2 Grupos no auspiciados por el W3C.....	68
2.4 Aplicaciones software para el desarrollo de la Web Semántica.....	77
2.5 Viabilidad del proyecto Web Semántica.....	88
2.6 La Web Semántica en el contexto de las bibliotecas digitales.....	95
2.6.1 Los CRAI: un marco de desarrollo para la Web Semántica .....	100

---



**Capítulo 3: Servicios de valor añadido en bibliotecas digitales**

3.1 Introducción.....	107
3.2 Sistemas de filtrado y recomendación de información.....	111
3.2.1 Criterios de clasificación.....	114
3.2.2 Sistemas de recomendación semánticos.....	117
3.3 Sindicación de contenidos.....	128
3.3.1 Características.....	130
3.3.2 Filtrado de canales RSS.....	132
3.4 Herramientas para el desarrollo de servicios de valor añadido.....	135
3.4.1 Sistemas de organización del conocimiento (KOS).....	135
3.4.1.1 Tesauros.....	141
3.4.1.2 Ontologías.....	144
3.4.1.3 Tesauros VS. Ontologías.....	146
3.4.2 Tecnologías de Web Semántica.....	147
3.4.2.1 XML.....	148
3.4.2.1.1 Familia de especificaciones.....	152
3.4.2.1.2 Sintaxis básica.....	156
3.4.2.1.3 Tecnologías auxiliares.....	159
3.4.2.2 RDF y RDFS.....	162
3.4.2.2.1 Sintaxis básica.....	165
3.4.2.3 FOAF.....	174
3.4.2.3.1 Sintaxis básica.....	174
3.4.2.4 RSS 1.0 (RDF Site Summary).....	181
3.4.2.4.1 Sintaxis básica.....	182
3.4.2.4.2 Módulos RSS 1.0.....	183
3.4.2.5 OWL.....	186
3.4.2.5.1 Sintaxis básica.....	187
3.4.3 Modelado lingüístico difuso.....	198
3.4.3.1 Introducción.....	198

---

3.4.3.2 Modelado lingüístico difuso clásico.....	199
3.4.3.2.1 Conjuntos difusos y funciones de pertenencia.....	200
3.4.3.2.2 Variables lingüísticas.....	201
3.4.3.2.3 Elección del conjunto de términos.....	202
3.4.3.3 Modelado lingüístico difuso ordinal.....	204
3.4.3.4 Modelado lingüístico difuso basado en 2-tuplas.....	209
3.4.3.5 Otras aproximaciones el modelado lingüístico.....	213

#### **Capítulo 4: Modelo de servicio semántico-difuso de DSI para bibliotecas digitales**

4.1 Introducción.....	215
4.2 Sistemas multiagente.....	217
4.3 Definición del modelo .....	221
4.3.1 Elementos.....	225
4.3.1.1 Tesauro.....	225
4.3.1.2 Perfiles de usuario.....	226
4.3.1.3 Documentos.....	227
4.3.1.4 Canales RSS.....	231
4.3.2 Módulos del sistema.....	235
4.3.2.1 Módulo de creación de perfiles y canales RSS.....	235
4.3.2.2 Módulo de <i>push</i> de información.....	238
4.3.2.3 Módulo de actualización de perfiles o <i>feedback</i> .....	245
4.3.2.2 Módulo de recomendación colaborativa.....	248

#### **Capítulo 5: D-FUSSION: Prototipo de servicio semántico-difuso de DSI para bibliotecas digitales universitarias**

5.1 Introducción.....	251
5.1.1 Nuevos enfoques de acceso a la información.....	251
5.2 Arquitectura de la aplicación.....	257
5.3 Estructura del prototipo.....	258

5.3.1 Elementos.....	259
5.3.2 Módulos.....	262
5.3.2.1 Módulo de <i>push</i> de informa.....	262
5.3.2.2 Módulo de recomendación colaborativa.....	265
5.3.2.3 Módulo de actualización de perfiles o <i>feedback</i> .....	267
5.4 Ejemplo operacional.....	269
<b>Capítulo 6: Resultados, conclusiones y trabajo futuro</b>	
6.1 Resultados.....	275
6.2 Conclusiones.....	276
6.3 Trabajos futuros.....	279
<b>Bibliografía</b>	283
<b>Referencias</b>	299
<b>Anexos</b>	309

---

---

# ÍNDICE DE FIGURAS Y TABLAS

## FIGURAS

---

Fig. 1 Distribución de fuentes según tipología.....	14
Fig. 2 Gráfica de cocitación de autores.....	16
Fig. 3 Gráfica de cocitación de autores más citados.....	20
Fig. 4 Niveles de la Web Semántica.....	35
Fig. 5 Ejemplo de documento XML.....	37
Fig. 6 Ejemplo de DTD.....	38
Fig. 7 Ejemplo de definición de un tipo de dato con XML Schema.....	39
Fig. 8 Grafo RDF y su representación con sintaxis XML.....	40
Fig. 9 Ejemplo de documento RDF.....	41
Fig. 10 Ejemplo de clasificación jerárquica.....	42
Fig. 11 Ejemplo de documento RDFS.....	42
Fig. 12 Ejemplo de consulta SPARQL.....	44
Fig. 13 Ejemplo de clase definida en OWL.....	46
Fig. 14 Ejemplo de regla definida en SWRL.....	48
Fig. 15 Esquema de componentes básicos de una URI.....	160
Fig. 16 Ejemplo del uso de espacios de nombre en XML.....	161
Fig. 17 Grafo RDF.....	163
Fig. 18 Representación del grafo de la Fig. 17 con sintaxis XML.....	164
Fig. 19 Declaración de elementos de clase.....	167
Fig. 20 Ejemplo de uso del elemento <i>rdf:parseType="Literal"</i> .....	170
Fig. 21 Ejemplo del uso de <i>rdf:value</i> .....	171
Fig. 22 Ejemplo de perfil de usuario en formato FOAF.....	179
Fig. 23 Unión, intersección y complemento de conjuntos.....	191
Fig. 24 Semántica asociada a un conjunto de 7 etiquetas lingüísticas.....	205

---

Fig. 25 Sistema multi-agente de recuperación y filtrado de información.....	221
Fig. 26 Niveles del modelo de servicio de DSI.....	224
Fig. 27 Ejemplo de ítem del canal RSS.....	234
Fig. 28 Proceso de creación de perfiles y canales RSS.....	238
Fig. 29 Solapamiento semántico de conceptos (I).....	241
Fig. 30 Solapamiento semántico de conceptos (II).....	242
Fig. 31 Proceso de generación de alertas.....	245
Fig. 32 Proceso de recomendación colaborativa.....	249
Fig. 33 Representación de la función $\mu$ .....	256
Fig. 34 Estructura básica del prototipo D-Fusion.....	259
Fig. 35 Interfaz de entrada de datos.....	263
Fig. 36 Relevancia de los resultados.....	264
Fig. 37 Recomendación colaborativa.....	266
Fig. 38 Menú desplegable de valoración.....	267
Fig. 39 Situación de conceptos en el tesauro.....	270

## TABLAS

---

Tabla 1. Clases y propiedades de FOAF.....	177
Tabla 2. Módulos de RSS 1.0.....	183
Tabla 3. Resultados del operador LOWA con $m=2$ y $\omega_1=0.4$ .....	208
Tabla 4. Documentos comunes en Biblioteconomía y Documentación.....	228
Tabla 5. Distancias.....	270
Tabla 6. Similaridades.....	271





---

# CAPÍTULO 1

## INTRODUCCIÓN

### 1.1 JUSTIFICACIÓN

El acceso universal a Internet ha revolucionado la forma en la que las personas acceden a los recursos de información que necesitan para desempeñar una tarea determinada o satisfacer una necesidad de información concreta (ya sea, por ejemplo, utilizando motores de búsqueda u ojeando directorios de páginas web). De hecho muchos de nuestros hábitos cotidianos se han visto modificados, pudiendo realizarlos desde nuestra propia casa utilizando cualquier terminal con acceso a la red (como leer el periódico, reservar un vuelo o presentar la declaración del impuesto sobre la renta).

La Web ha simplificado enormemente el acceso a la información, así como la publicación y redifusión de contenidos, pero como contrapartida cada vez se hace más complicado para los usuarios de Internet recuperar recursos relevantes que satisfagan sus necesidades de información (Lawrence, Giles 1999).

Sin embargo, a pesar de que la introducción de las nuevas tecnologías de la información en nuestra vida ha sido muy rápida, gracias principalmente a su facilidad de uso y transparencia para el usuario (no podemos obviar lo sencillo que resulta para cualquier persona navegar por Internet o utilizar un buscador), comprobamos que aún están lejos de ofrecer un servicio tan eficiente y completo como cabría esperar.

Valga como ejemplo una simple consulta en uno de los muchos buscadores que podemos encontrar en la red. Cuando introducimos una serie de palabras clave, su motor de búsqueda se limita a aplicar a estas un algoritmo que mediante técnicas estadísticas nos devuelve un listado de documentos ordenados por

---



relevancia. Sin embargo, generalmente este listado de resultados no satisface adecuadamente la necesidad de información que generó la consulta al sistema.

Esto sucede porque el motor de búsqueda es incapaz de “*comprender*” qué pretendía buscar exactamente el usuario. Para el buscador las palabras clave no son más que cadenas de caracteres completamente descontextualizadas y carentes de un significado real, que presentan una cierta frecuencia de aparición entre los documentos que están indizados en el sistema. Por lo tanto, básicamente se limita a recuperar aquellos recursos en los que esa frecuencia de aparición es mayor (aunque ya existen algoritmos más complejos como *PageRank* [1] que además toman en consideración otros factores como, por ejemplo, el número de enlaces que recibe un documento desde otros documentos).

Podríamos afirmar, por lo tanto, que para el motor de búsqueda es lo mismo un documento sobre los *bancos* de atunes del Atlántico, que una noticia sobre el *Banco* Mundial, o que el sitio web de una fábrica de *bancos* de madera.

La brecha entre lo que el usuario demanda y los recursos recuperados se debe, principalmente, a que el sistema necesita trabajar con representaciones tanto de los documentos como de las necesidades de información de los usuarios para poder operar con ellos en el ámbito de uno de los diferentes modelos teóricos que abordan el problema de la recuperación de información (básicamente el modelo booleano, el vectorial y el probabilístico).

Por un lado, el contenido semántico de los documentos se suele representar mediante índices de frecuencias de aparición de términos, mientras que las necesidades de información suelen ser expresadas por los propios usuarios utilizando una serie de términos de búsqueda (en muchos casos ambiguos o inapropiados -(Kristensen 1999) (McCray, Srinivasan, Browne 1994)-). Este problema adquiere una mayor dimensión cuando el usuario presenta un alto grado de especialización y requiere recursos muy específicos (como es el caso de

---

los investigadores y su necesidad de acceder a publicaciones científicas de su área de interés (Bollacker, Lawrence, Giles 2000)).

A esta transformación en la forma de acceso a los recursos informativos no se han visto ajenas las bibliotecas. Al igual que la propia Web, deben hacer frente al reto de gestionar eficazmente el gran volumen de documentos que almacenan para facilitar a sus usuarios un acceso sencillo y ágil a recursos que satisfagan sus necesidades de información. En las bibliotecas, tradicionalmente se han propuesto diferentes soluciones como, por ejemplo, los servicios de difusión selectiva de información (DSI), los boletines de novedades o los boletines de sumarios.

La progresiva penetración de las nuevas tecnologías de la información y la comunicación (TIC) en las bibliotecas, la universalización del uso de Internet y la diversificación de los recursos que se pueden hacer accesibles desde la red ha provocado que las bibliotecas se embarquen en un proceso de reinención que implica una profunda revisión de sus técnicas y metodologías de trabajo, y de los servicios que prestan para tratar de adaptarlos a las nuevas herramientas tecnológicas que tienen a su disposición, y así de esta forma poder satisfacer mejor las demandas cada vez más exigentes y específicas de sus usuarios.

Como fruto de este proceso de transformación y adaptación surgen las bibliotecas digitales, una extensión lógica de las bibliotecas físicas (Marchionini 2000) que albergan colecciones de recursos en formato electrónico (bien en su origen, bien tras ser sometidos a procesos de digitalización) y que tienen asociados una serie de servicios para facilitar el acceso a estos recursos a diferentes comunidades de usuarios utilizando para ello diversas tecnologías (Smeaton, Callan 2005).

Las bibliotecas digitales son sistemas híbridos que heredan muchas de las virtudes e inconvenientes de las bibliotecas físicas, ya que se ha producido un trasvase a la Web de gran parte de sus herramientas y servicios, aunque adaptados a la nueva plataforma.

---

Este ha sido el caso de los servicios de DSI, los que pueden considerarse como los sistemas de filtrado y recomendación de información de las bibliotecas. Estos servicios, de acuerdo al perfil de los usuarios suscritos al servicio, generan periódicamente (o a petición del propio usuario) una serie de alertas en los que se les notifica qué recursos existen en la biblioteca que se adecuan a sus intereses (Aksoy et al. 1998) (Foltz, Dumais 1992).

El funcionamiento de los servicios de DSI es, de hecho, muy similar al de los sistemas de filtrado y recomendación. Estos sistemas se basan en la aplicación de diferentes técnicas de filtrado que tratan de gestionar una serie de procesos que pretenden proporcionar a la gente exclusivamente la información que necesitan o que puede ser de su interés. En dominios textuales los sistemas de filtrado o de recomendación evalúan y filtran recursos disponibles en la Web u otros sistemas de información (normalmente en formato XML o HTML) para asistir a las personas en sus procesos de búsqueda de información (Resnick, Varian 1997), en muchos casos mediante sistemas multi-agente.

Tradicionalmente estos sistemas se clasifican en dos categorías básicas (Popescul et al. 2001): sistemas de filtrado basados en contenido y sistemas de filtrado cooperativo. La principal diferencia entre ambos radica básicamente en el uso de la opinión de uno o varios usuarios para realizar el filtrado de recursos.

Los sistemas de recomendación basados en contenido filtran la información y generan recomendaciones equiparando una serie de palabras clave definidas por el usuario con los términos usados para representar el contenido de los documentos, ignorando cualquier información proporcionada por otros usuarios. Estos sistemas suelen ser poco eficientes cuando se conoce poco sobre las necesidades de información de cada individuo, por lo que es frecuente recurrir al uso de perfiles que caracterizan a los usuarios a través de sus preferencias explícitas o implícitas. De hecho la integración de los perfiles de usuario en los procesos de filtrado suponen una potente herramienta para proporcionar un filtrado rápido y eficiente (Shapira et al. 1997).

---

Por su parte, los sistemas de filtrado cooperativo utilizan la información proporcionada por diversos usuarios para recomendar documentos a un usuario dado, ignorando la representación que se haga de su contenido. Es bastante usual agrupar a los usuarios en diferentes categorías o estereotipos que están caracterizados por una serie de reglas y preferencias definidas por defecto que representan las necesidades de información y hábitos comunes a un grupo de usuarios relacionados. Sin embargo estos sistemas suelen ser poco eficientes cuando se dispone de poca información sobre el usuario o sus intereses son poco habituales.

La tendencia actual es desarrollar sistemas híbridos que aprovechen las ventajas que presentan ambos sistemas.

No obstante, estos servicios adolecen de algunas carencias:

- los procesos de comunicación entre agentes, y entre agentes y usuarios se ven dificultados por las diversas formas en que la información está representada;
- esta heterogeneidad en la representación de la información provoca a su vez que esta no pueda ser reutilizada en otros procesos y por otras aplicaciones;

Una posible solución puede consistir en mejorar tanto la representación de las necesidades de los usuarios, como la representación de la información disponible en la biblioteca digital, de forma que esta sea lo más homogénea y comprensible posible tanto para los usuarios como para los agentes. A continuación comentamos qué técnicas y tecnologías podrían permitirnos cumplir estos objetivos.

#### *a) Representación de las necesidades de los usuarios*

En los sistemas de recuperación de información tradicionales son los propios usuarios los que definen sus necesidades de información en forma de consultas

---

que componen con uno o varios términos de búsqueda y opcionalmente utilizando una serie de operadores.

Sin embargo, esto implica que en muchas ocasiones estas necesidades estén expresadas de una forma subjetiva, imprecisa o vaga (Zadeh 1975), ya que la pertinencia de la consulta depende en gran medida de la destreza del propio usuario para definirla. Por lo tanto se hace necesario el desarrollo de sistemas auxiliares en los que el usuario, ya sea adoptando una postura activa o pasiva, pueda obtener unos resultados más precisos y exhaustivos.

Si se opta porque el usuario defina él mismo sus consultas se deben poner a su disposición herramientas que le faciliten el proceso de recuperación. Así, por ejemplo, la aplicación técnicas de inteligencia artificial como la lingüística difusa (Herrera-Viedma 2001) puede ser una solución viable, ya que proporciona un modelo de representación de la información flexible (fácil de interpretar por los usuarios y con la que pueden operar las máquinas), que permite procesar información cualitativa mediante el uso de etiquetas lingüísticas. Otras soluciones proponen el desarrollo de técnicas para el enriquecimiento de las consultas formuladas y la desambiguación de los términos de búsqueda definidos por los usuarios utilizando elementos auxiliares como tesauros u ontologías (Kristensen 1993) (Paice 1991).

Si se opta por definir una postura pasiva del usuario la mejor solución es desarrollar sistemas de filtrado de información. Estos sistemas permiten evaluar y seleccionar la información que mejor se adapta a las necesidades de los usuarios finales, facilitándoles el acceso, exclusivamente, a recursos de su interés y reduciendo de esta forma el ruido documental.

Su principal objetivo es dar respuesta a las necesidades de los usuarios más persistentes en el tiempo, generando recomendaciones de información personalizadas (de ahí que se haya popularizado la denominación de algunos de estos sistemas como sistemas de recomendación -(Resnick, Varian 1997)-). Para ello, en lugar de representar las necesidades de información mediante consultas

---

puntuales definidas en el momento de la búsqueda, el sistema dispone a priori de información sobre cada individuo, que viene representada en forma de perfiles de usuario, a partir de los cuales se generan una serie de consultas pasivas.

Estos perfiles no son más que representaciones de las preferencias, intereses y características de cada individuo, que son recopiladas mediante métodos explícitos y/o implícitos (bien interrogando directamente al individuo, bien monitorizando su comportamiento al interactuar con el sistema).

#### *b) Representación de la información*

Como hemos visto anteriormente, los modelos tradicionales de recuperación de información se basan en la equiparación léxica de términos, lo cual les impide, por ejemplo, recuperar diferentes recursos que, aún siendo relevantes para una necesidad de información concreta, utilizan diferentes términos para definir un mismo concepto (términos sinónimos).

Esta forma de representar la información como una mera sucesión de cadenas de caracteres restringe la capacidad de recuperar información relevante en el modelo de Web actual. Una posible solución para solventar estas deficiencias consistiría en mejorar y enriquecer la representación de la información utilizando un vocabulario y modelo de datos común que fuera interpretable tanto por humanos como por los propios agentes.

Precisamente en esta idea se fundamenta el proyecto Web Semántica (Berners-Lee, Hendler, Lassila 2001), una extensión de la Web actual cuyo objetivo último consiste en convertirse en una plataforma universal para el intercambio de información. Se apoya básicamente en dos ideas: el marcado semántico de los recursos (lo que implica una separación formal entre el contenido y la estructura de los documentos), y el desarrollo de agentes software capaces de procesar y operar con estos recursos a nivel semántico (Berners-Lee 2000) (Hendler 2001).

---

Además, esta representación estandarizada de la información acaba redundando en una mejora significativa de la calidad de la información recuperada por los motores de búsqueda, y permite, además, preservar tanto la independencia de plataforma, como fomentar el intercambio y reutilización de contenidos.

La propuesta del W3C para el etiquetado semántico de recursos es RDF (*Resource Description Framework*) (Becket 2004), un metalenguaje y que proporciona un modelo de datos que permite estructurar la el conocimiento en forma de tripletas sujeto-predicado-objeto que a su vez se pueden representar en forma de grafos de nodos y arcos que están identificados unívocamente mediante referencias URI (*Uniform Resource Identifiers*), o identificadores uniformes de recursos.

Esta primera etapa de marcado de recursos es crucial para el éxito del modelo ya que sin la existencia de una masa crítica suficiente de documentos etiquetados no será posible su implantación.

El siguiente paso consistiría en desarrollar agentes inteligentes, unas entidades software que realizan de forma semiautónoma tareas de recogida, filtrado, y procesamiento de información e inferencia, y que cumplen el papel de «infomediarios» entre las necesidades de información de los individuos y las fuentes de recursos distribuidas de la red. De hecho, el verdadero potencial de la Web Semántica se desarrollará cuando la web esté poblada por un gran número de agentes que sean capaces de recoger información de diferentes fuentes, procesarla e intercambiar resultados con otros agentes.

La idea de reutilización o interoperabilidad de recursos es la base del modelo de Web Semántica definido por Berners-Lee (2000), que gráficamente se representa mediante un modelo en el que se definen diferentes niveles de desarrollo:

- en los tres primeros se establecen las bases para poder representar la información de una manera semánticamente accesible, definiendo una sintaxis y un modelo de datos común (RDF/XML) y un vocabulario que
-

permite crear ontologías web. permiten a los agentes inteligentes realizar razonamientos y obtener inferencias sobre los recursos disponibles en la red. Existen diferentes lenguajes para diseñar ontologías web (Pulido et al. 2006), aunque la apuesta del W3C es OWL (*Ontology Web Language*) (McGuinness, van Harmelen 2004), un lenguaje de gran capacidad expresiva que mantiene la convención sintáctica de RDF/XML.

- en las capas superiores se definen los elementos que permiten a los agentes software autenticar y comprobar la confiabilidad de los diferentes elementos del modelo (recursos, agentes, inferencias obtenidas, etc.) gracias al desarrollo de redes de confianza y al uso de firmas digitales.

El modelo procura la máxima interoperabilidad de cada capa con las capas adyacentes lo que implica que los recursos definidos en una capa determinada pueden ser reutilizados íntegramente por las capas superiores, pero sólo parcialmente por las inferiores.

Como vemos, la combinación los sistemas de información y las tecnologías de Web Semántica se presenta como una solución prometedora que permitirá mejorar la eficiencia de los sistemas de acceso a la información, así como resolver problemas endémicos presentes en todos los procesos de recuperación de información y abrir la puerta al desarrollo de nuevos servicios que permitan satisfacer necesidades de información muy específicas o diferentes de las habituales.

---



## 1.2 OBJETIVOS

El objetivo general de este trabajo persigue la mejora del acceso a recursos informativos de interés a los usuarios de las bibliotecas digitales universitarias. Para conseguirlo, los objetivos específicos planteados son:

- la mejora de la representación de la información y de la comunicación entre usuario y sistema utilizando para ello tecnologías de Web Semántica y técnicas de modelado lingüístico difuso,
- la mejora de la satisfacción de las necesidades de información del usuario mediante el filtrado de recursos de acuerdo a las preferencias e intereses recogidos en su perfil,
- y el desarrollo de nuevos servicios que permitan satisfacer necesidades de información específicas de la comunidad de usuarios a la que sirven las bibliotecas digitales universitarias

En este trabajo queremos profundizar en el diseño de sistemas de acceso y recomendación de información, realizando una propuesta de mejora basada en la hibridación de sistemas semánticos de recuperación y filtrado (que son complementados con la aplicación de otras técnicas de *Soft Computing* como el modelado lingüístico difuso).

Con el desarrollo de este modelo pretendemos demostrar que la aplicación de un sistema de dichas características en un dominio específico puede ofrecer a los usuarios información relevante y personalizada, y de esta forma incrementar su satisfacción.

---

## 1.3 METODOLOGÍA

A continuación procedemos a analizar las principales herramientas y materiales empleados en el desarrollo de la investigación.

### 1.3.1 INVESTIGACIÓN DE CAMPO

Se ha desarrollado un prototipo de sistema de difusión selectiva de información (DSI) basado en tecnologías de Web Semántica y técnicas de modelado lingüístico difuso para comprobar la viabilidad del modelo propuesto en la memoria.

El prototipo se ha desarrollado casi íntegramente en PHP 5, utilizando las librerías básicas del lenguaje y la librería DOM. La única salvedad es una pequeña aplicación programada en JavaScript que permite desplegar el nombre de los usuarios cuyas valoraciones han sido utilizadas para generar las recomendaciones colaborativas del servicio de DSI.

Para ejecutar el prototipo de manera local ha sido necesario instalar un servidor Apache 2 con soporte para PHP 5.

Adicionalmente se han utilizado una serie de lenguajes de etiquetado semántico como RDF Schema, RSS y FOAF para definir respectivamente el tesoro, los canales RSS y los perfiles de usuario con los que opera la aplicación.

### 1.3.2 INVESTIGACIÓN DOCUMENTAL

Las fuentes de información utilizadas en este trabajo han sido muy diversas.

1. *Obras de referencia*: Básicamente diccionarios monolingües generales como el de la Real Academia de la Lengua [2], diccionarios bilingües generales como el Diccionario Collins (Smith 1997) o WordReference [3], diccionarios especializados en Ciencias de la Documentación [4], y diccionarios y glosarios específicos del área de las Ciencias de la Computación e Internet, como el
-

glosario de términos de Internet [5]. También se han utilizado diferentes manuales y tutoriales en línea de los lenguajes de programación necesarios para el desarrollo de la aplicación resultado del trabajo de investigación.

2. *Monografías y revistas científicas impresas*: Tanto las monografías como las revistas científicas en formato papel se han obtenido de la Biblioteca General Universitaria de la Universidad de Granada [6]. No obstante, su uso ha sido bastante residual ya que la gran mayoría de las fuentes consultadas estaban en formato electrónico.
  3. *Bases de datos científicas en línea*: Principalmente se han utilizado diferentes bases de datos en línea sobre literatura científica que ofrecen acceso a recursos en formato electrónico a texto completo.
    - *Web of knowledge* [7]: Es una base de datos científica que cubre 250 disciplinas de las áreas de las ciencias puras y aplicadas, las ciencias sociales, y las artes y humanidades. Contiene alrededor de 8.500 revistas revisadas por pares, que suponen en total un 1,3 millones de artículos; además incluye cerca de 2.000 monografías y más de 400.000 recursos web a texto completo.
    - *ScienceDirect* [8]: Esta base de datos dispone del 25% de la literatura científica, tecnológica y médica mundial a texto completo. Incluye libros electrónicos, obras de referencia, manuales, series de monografías y casi 8 millones de artículos.
    - *ACM Portal* [9]: El portal dispone de una biblioteca digital que contiene una gran colección de recursos a texto completo de revistas científicas, memorias, boletines de noticias y actas de congresos especializados en Ciencias de la Computación.
    - *Scopus* [10]: Es una base de datos de resúmenes y citas sobre literatura científica y recursos web. Por poner solo un ejemplo, dispone de 33 millones de resúmenes entre artículos de revistas revisadas por pares y
-

de acceso libre. Además proporciona enlaces a los sitios web de los proveedores donde se puede acceder al texto completo (si se es suscriptor).

- *E-LIS (Open archive for Library and Information Science)* [11] es un repositorio de libre acceso que contiene más de 7.000 documentos científicos y técnicos especializados en el área de la Biblioteconomía y la Documentación (así como otras áreas afines), y que son cedidos al repositorio por la propia comunidad de usuarios. Este repositorio ha servido, además, como fuente principal para obtener las descripciones de los recursos que han sido utilizadas en el desarrollo de la aplicación presentada en este trabajo.

4. *Recursos web*: Otros muchos recursos (que representan más de la mitad del total) relativos a tecnologías, especificaciones o proyectos relacionados con la Web Semántica, se han hallado realizando consultas a motores de búsqueda, o acudiendo directamente a los sitios web de organismos e instituciones implicados en su desarrollo. La principal fuente de información ha sido la página del World Wide Web Consortium (W3C), donde se encuentra la información más actualizada y la mayoría de las especificaciones de los vocabularios utilizados en este trabajo. El resto de especificaciones no auspiciadas por el W3C se han consultado en sus respectivos sitios web. En menor medida también se han utilizado otro tipo de recursos como presentaciones creadas para talleres o comunicaciones a congresos, e incluso algunas noticias y comentarios publicados en *blogs* especializados en el dominio, que en los últimos años se están convirtiendo en canales dinámicos de difusión e intercambio de ideas.
-

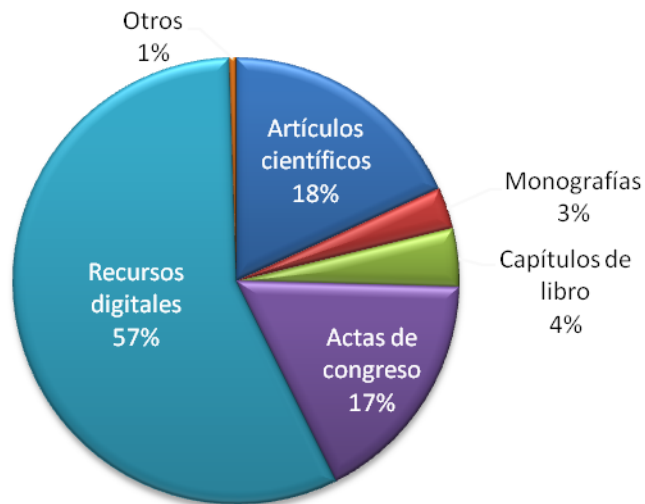


Fig.1 Distribución de fuentes según tipología

### 1.3.2.1 ANÁLISIS DE COCITACIÓN DE AUTORES

Para completar el estudio de las fuentes utilizadas en esta memoria se ha recurrido al uso de técnicas de análisis de redes sociales (White 2003) (White, McCain 1998) para visualizar gráficamente las áreas temáticas que cubren nuestro estudio gracias al estudio de la cocitación de autores, que permiten determinar y representar las relaciones sociales e intelectuales implícitas de un autor con otros autores a los que cita en sus trabajos, de los que recibe citas, o con los que comparte la autoría de un trabajo.

Se han realizado dos análisis de cocitación: uno sobre el total de autores y otro sobre los autores más citados, utilizando en ambos casos la misma metodología. A partir de la bibliografía utilizada en esta memoria se obtienen los registros de aquellos recursos que aparecen indizados en la base de datos *ISI Web of Knowledge*.

De estos registros se extrae el listado de autores cuyos trabajos aparecen referenciados en el campo "referencias citadas" (*cited references*). Posteriormente se tratan de manera adecuada para poder ser procesados por una aplicación específica para la representación de redes sociales [12].

Para el primero de los análisis de cocitación que realizamos utilizamos el listado del total de autores recuperado (alrededor de 900). Dado que trabajar con tal cantidad de autores acaba generando gráficos muy abigarrados (confusos y difícilmente interpretables) se ha sometido a la matriz original a una poda de datos estableciendo restricciones sobre el número conexiones por nodo (se han admitido solo aquellas que tienen un grado mayor o igual a 1), y sobre la intensidad de estas conexiones (definiendo un número de cocitaciones entre autores mayor o igual a 12).

Para refinar aún más los datos, al conjunto de autores resultante se le ha aplicado el algoritmo de poda *pfnet (Pathfinder)* (White 2003) obteniéndose como resultado un total de 60 autores de entre los casi 900 que aparecían en la matriz original. De esta manera el gráfico resultante gana en claridad, aunque sea a costa de una pérdida de información. El resultado de este proceso es el que comentamos a continuación.

A primera vista, si trazamos una bisectriz imaginaria (desde el ángulo superior derecho del gráfico al inferior izquierdo) comprobamos que es posible dividir el gráfico en dos grandes grupos de autores bien diferenciados. Por debajo de la bisectriz encontramos a los autores relacionados con la recuperación de información, y por encima de ella los autores especializados en lógica difusa.

---

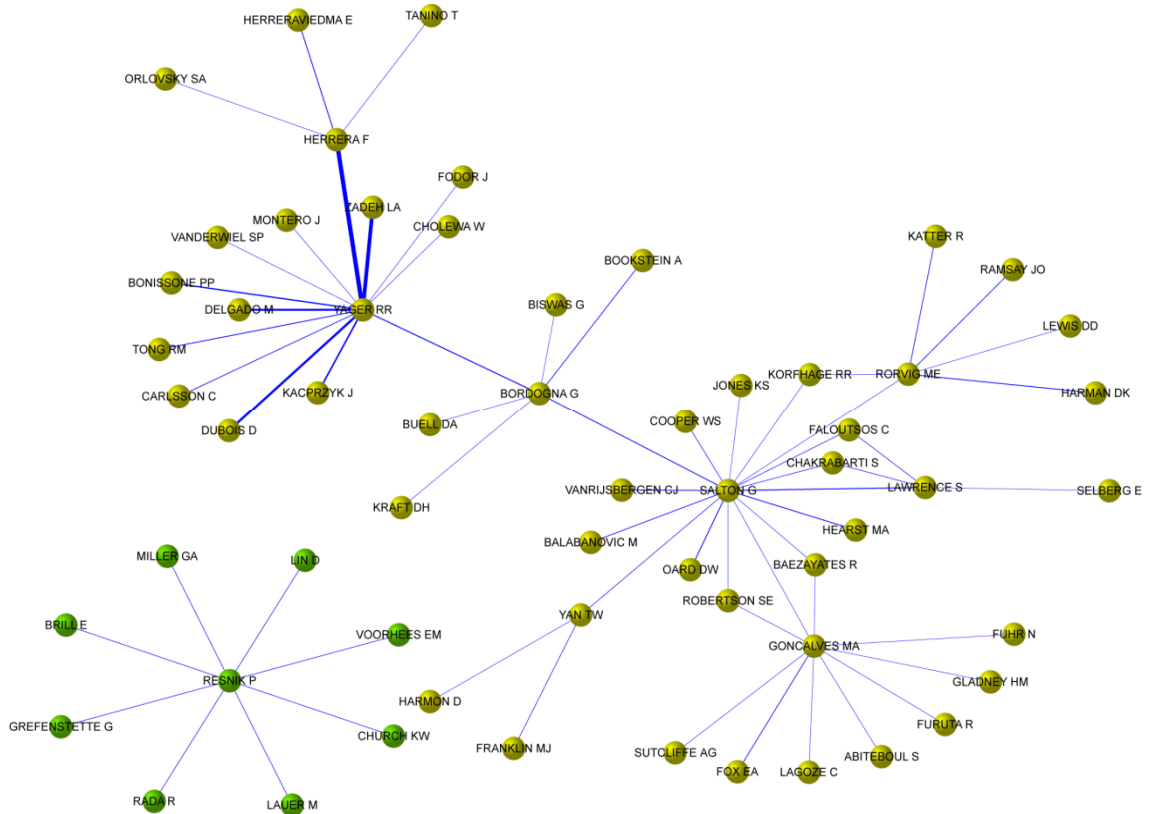


Fig.2 Gráfico de cocitación de autores

Generalmente en este tipo de representaciones la posición central la suelen ocupar áreas de conocimiento multidisciplinares. Sin embargo en este caso, y debido a la evidente polarización entre los dos grupos principales, la posición central aparece ocupada por el autor que sirve de puente entre ambos.

En la parte inferior derecha podemos identificar un grupo de autores bastante amplio y ramificado en el que es posible distinguir hasta 5 áreas de conocimiento. La principal de ellas sería la que ocupa la posición central del grupo (por lo tanto, la más interdisciplinar) y que está formada por autores especializados en *recuperación de información*. Este dominio se forma entorno a un único autor (G. Salton, el padre de la teoría de la recuperación de información moderna) que ejerce de nexo común entre la mayoría de los autores de su área (relacionando al resto de dominios que es posible distinguir dentro de este grupo

de autores), y que sirve como elemento conectivo con el otro gran dominio presente en el gráfico.

Por encima del grupo principal y a la derecha se define el dominio de *procesamiento de información y análisis de datos*, donde M.E. Rorvig representa el papel de *broker* con el área de *recuperación de información* (estableciendo relación con Salton y con R.R. Korfhage, autor especializado en la recuperación de información mediante el uso de interfaces gráficos).

Justo debajo de esta área es posible distinguir un dominio algo más reducido formado únicamente por el autor E. Selberg, que está especializado en la *recuperación de información en la Web* y que por sí solo podría constituir un área subordinada al grupo principal (con el que conecta a través de S. Lawrence).

El siguiente dominio, situado en la esquina inferior izquierda del gráfico, es el más extenso (por el número de autores que aglutina) de los directamente relacionados con la *recuperación de información*, y se corresponde con el área de las *bibliotecas digitales*. Aquí encontramos destacados autores como S.E. Robertson, M.A. Gonçalves y R. Baeza-Yates, que entroncan directamente con el área de recuperación de información a través de Salton.

A la izquierda del grupo principal encontramos el cúmulo de autores especializados en el *filtrado de información*, y que entronca directamente con el área de *recuperación de información* a través de T.W. Yan.

---



El último dominio presente en esta parte inferior del gráfico es el que se corresponde con los investigadores especializados en el *procesamiento del lenguaje natural* (PLN) y que aparecen aislados<sup>1</sup> en un grupo donde los autores se disponen en forma de estrella entorno a P. Resnick. Este aislamiento puede deberse a varios factores (absoluta independencia entre las áreas de investigación, malas relaciones de este grupo de investigadores con el resto de la comunidad científica, etc). No obstante, nos inclinamos a pensar que esta situación se debe a las estrictas restricciones impuestas en la poda de datos (punto que se confirma a la vista del análisis de los datos de cocitación de los autores más citados que se realiza más adelante).

Por encima de la bisectriz imaginaria que hemos trazado en el gráfico se define el otro gran grupo de autores donde se agregan los investigadores especializados en *conjuntos difusos*. Aparecen un total de 16 autores de entre los cuales destacan R.R. Yager, que cumple el papel de *broker* del grupo, L.A. Zadeh (padre de la teoría de conjuntos difusos), y F. Herrera que aglutina conexiones con otros autores especializados en el desarrollo de sistemas de toma de decisiones (como E. Herrera-Viedma).

El enlace natural entre el grupo de *recuperación de información* y el de *conjuntos difusos* es G. Bordogna, un autor especializado en la recuperación de información en sistemas difusos. También relacionados con este autor encontramos a otros investigadores que trabajan sobre el desarrollo de sistemas de recuperación basados en pesos, como D.H. Kraft, o que realizan análisis bibliométricos, como A. Bookstein.

---

<sup>1</sup> Es preciso indicar que la posición que ocupa en el gráfico el dominio de *procesamiento del lenguaje natural* ha sido modificada (respetando la disposición original de los autores) para obtener una mayor claridad y comprensibilidad del mismo. La posición original del grupo prácticamente se superpone a la que ocupa el área de *recuperación de información* (apareciendo P. Resnik muy próximo a G. Salton).

---

No obstante, al analizar este primer gráfico nos resulta muy llamativo que no aparezca un área específica para el dominio de la Web Semántica, a pesar de que el grueso de este trabajo se basa en sus tecnologías, y de que entre los autores más citados aparecen muchos que son clave en el desarrollo del proyecto (como, por ejemplo, T. Berners-Lee).

Una posible explicación podría encontrarse en la relativa juventud del dominio (no olvidemos que el trabajo sobre el que se fundamenta el modelo de Web Semántica data de 2001). Esto provoca que, a pesar de que el área se define apoyándose en múltiples disciplinas (recuperación y filtrado de información, bibliotecas digitales, etc.) la intensidad de las relaciones que establece con estas es aún muy débil, y como consecuencia dichas conexiones no se hacen evidentes cuando se realiza un análisis de redes sociales con restricciones de poda muy estrictas (como es el caso que nos ocupa). Por decirlo de otra manera, los autores especializados en Web Semántica citan y se basan en trabajos de autores de otras áreas, pero los autores de estos otros dominios aún no se hacen eco de las investigaciones llevadas a cabo en el dominio de la Web Semántica, y el nivel de colaboración e intercambio de ideas entre ellos es mínimo. De este razonamiento se podría concluir que la “comunidad semántica” se encuentra aún en proceso de maduración.

Ahora bien, esta situación no se da cuando lo que analizamos es el índice de cocitación entre los autores más citados (aquellos que han recibido 3 o más citas). El preprocesamiento del conjunto de autores (un total de 67) se ha limitado a la aplicación del algoritmo de poda *pfnet*, ya que al ser relativamente pocos no ha sido necesario imponer restricciones de grado e intensidad. Como resultado se ha obtenido el siguiente gráfico.

---

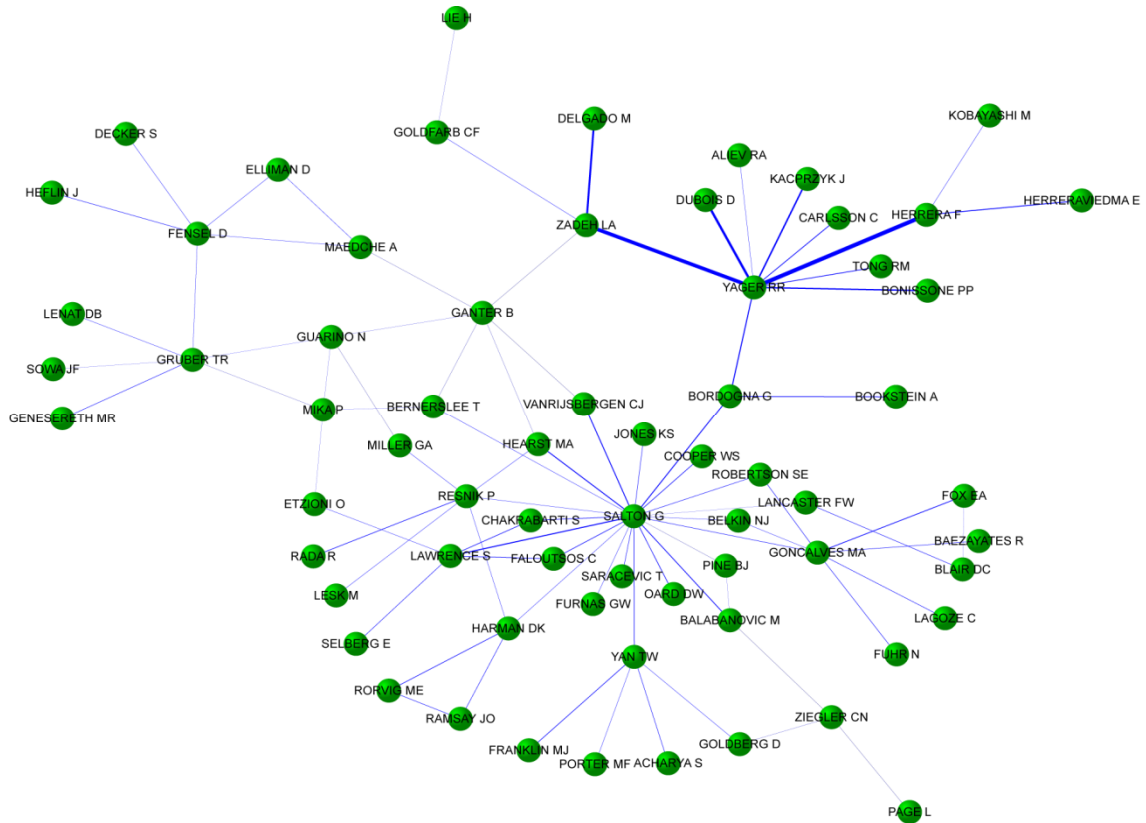


Fig.3 Gráfico de cocitación de autores más citados

Vemos que el gráfico resultante es bastante similar al anterior. En este caso la posición central del gráfico la ocupa el área de *recuperación de información* (lo cual da una idea de su nivel de interdisciplinariedad), y de nuevo Salton aparece como *acumulador* de enlaces y nexo entre las diferentes áreas que se definen en el gráfico.

Vuelven a distinguirse claramente el dominio de las *bibliotecas digitales* (donde aparecen nuevas figuras tan relevantes como Lancaster o Belkin), el de *filtrado de información*, el de *procesamiento de información y análisis de datos*, el de *recuperación de información en la Web*, el de *procesamiento del lenguaje natural* (esta vez con una clara conexión entre P. Resnik y G. Salton), y el de *conjuntos difusos*. Sin embargo, en este gráfico sí que aparecen un par de áreas que quedaban ocultas en el anterior: el dominio de *sistemas de recomendación* y el dominio de *sistemas basados en conocimiento*.

El dominio de *sistemas de recomendación* se encuadra entre el área de *bibliotecas digitales* y el de *filtrado de información*, y está formado por un reducido grupo de autores entre los que se encuentran C.N. Ziegler, L. Page y M. Balabanovic (que enlaza directamente con G. Salton).

Por su parte, el dominio de *sistemas basados en conocimiento* abarca un área extensa de la parte izquierda del gráfico. El autor que cumple el papel de *broker* con el dominio de *recuperación de información* es T. Berners-Lee, el conocido precursor de Internet y del proyecto Web Semántica.

Dentro de este dominio genérico se distinguen otros dos sub-dominios más específicos. Por un lado, en el ramal superior encontramos el que forman los autores especializados en Web Semántica (sistemas de conocimiento basados en ontologías) como A. Maedche, D. Fensel o B. Ganter (el cual establece una interesante conexión entre este dominio y el de lógica difusa a través de la relación que establecen con L. Zadeh).

En la otra rama encontramos autores como P. Mika, T. Gruber o M. Genesereth que centran su investigación sobre el desarrollo de redes semánticas, sistemas basados en conocimiento e inteligencia artificial. La conexión con el dominio de *recuperación de información* se realiza a través de O. Etzioni.

Ambas ramas se conectan entre sí a través de N. Guarino, especialista en el estudio de las bases ontológicas de la ingeniería del conocimiento y el modelado conceptual.

A la luz de los resultados obtenidos en el análisis de la bibliografía utilizada podemos concluir que los recursos y autores seleccionados hacen una cobertura adecuada de las diferentes áreas temáticas necesarias para definir las bases teóricas y metodológicas del trabajo desarrollado.

---

## **1.4 ESTRUCTURA DEL TRABAJO**

Esta memoria se desarrolla en seis capítulos: este primer capítulo de introducción, cuatro capítulos (del 2 al 5) en los que se describe la investigación realizada y un capítulo final en el que se discuten los resultados, y se apuntan una serie de conclusiones y trabajos futuros.

Para explicitar el contexto de investigación, en el capítulo 2 analizamos la Web Semántica como plataforma de acceso universal a la información en formato digital y la estrecha interrelación existente entre este nuevo paradigma tecnológico y las bibliotecas digitales. Analizamos también el papel clave que van a cumplir las bibliotecas digitales en el nuevo Espacio Europeo de Educación Superior como nodos de entrada a las redes de intercambio de información científica.

Para describir las bases instrumentales de la investigación, en el capítulo 3 se estudian algunos servicios de valor añadido que permiten mejorar la satisfacción de los usuarios de las bibliotecas digitales (como los sistemas de filtrado y recomendación de información, y los servicios de redifusión de contenidos), así como algunas herramientas que permiten desarrollar este tipo de servicios (como los sistemas de organización del conocimiento, las tecnologías del proyecto Web Semántica y las técnicas de modelado lingüístico difuso).

En el cuarto capítulo se presenta un modelo de servicio semántico-difuso de difusión selectiva de información (DSI) al que se han incorporado técnicas de filtrado mejoradas mediante la aplicación de técnicas lingüísticas difusas, que permiten la actualización dinámica de los perfiles de usuario basándose en la bonificación/penalización de preferencias.

---

En el capítulo 5 se procede a describir el prototipo D-FUSSION, un servicio de DSI semántico, mejorado con técnicas de modelado lingüístico difuso, diseñado para facilitar a investigadores del dominio de la Biblioteconomía y la Documentación el acceso a recursos de interés.

En el capítulo 6 se discuten los principales resultados obtenidos, se presentan las conclusiones del trabajo y se apuntan futuras líneas de investigación.

---



# WEB SEMÁNTICA Y BIBLIOTECAS DIGITALES

### 2.1 INTRODUCCIÓN

Hoy día, es un hecho constatable que estamos inmersos en la sociedad de la información. Lo queramos o no, muchos de nuestros hábitos más cotidianos nos remiten al uso de los nuevos sistemas y tecnologías de la información. El acceso a Internet se ha convertido en algo habitual y casi rutinario para leer el periódico, reservar vuelos o habitaciones de hotel, presentar la declaración del impuesto sobre la renta, etc. A modo de eslogan publicitario podríamos decir que tenemos *'todo un mundo de información a un clic de distancia'*.

Sin embargo, este modelo de Web se está colapsando y las herramientas de las que dispone ya no son capaces de gestionar de forma eficiente el acceso a la avalancha de información que circula por la red y que, además, no cesa de crecer exponencialmente. Coloquialmente podríamos afirmar que *"Internet está muriendo de éxito"*.

Pero por extraño que parezca, el problema del acceso a la información no es nuevo. Ya en 1945 el ingeniero y científico estadounidense Vannevar Bush en su artículo titulado *"As we may think"* (*Como podríamos pensar*) (Bush 1945) apunta que uno de los principales problemas que encuentran los científicos para desarrollar su trabajo es la dificultad para acceder al cada vez mayor número de recursos científicos, que no para de crecer exponencialmente. Un crecimiento que, según sus propias palabras, va *"mucho más allá de nuestra capacidad de aprovechar realmente la información acumulada. El conjunto de la experiencia humana está creciendo a un ritmo prodigioso, pero los medios que empleamos para desplazarnos por este laberinto hasta llegar al punto importante del momento son los mismos que utilizábamos en los tiempos de las carabelas"*.

---



Es decir, no solo el volumen de recursos se hace inabarcable, sino que además no se disponen de los medios adecuados para acceder a la información de interés de una manera rápida y eficiente.

Bush achaca esta ineficiencia en el acceso a la información no solo a la falta de medios técnicos adecuados, sino también a la incapacidad de los métodos de indización y clasificación tradicionales para reproducir el mecanismo que utiliza la mente humana para acceder a la información mediante la asociación de conceptos.

Como solución a este problema propone la creación de una hipotética máquina, denominada *Memex (Memory Extender)* que funciona a modo de extensión de memoria del usuario, y que utilizando los medios tecnológicos de la época (como los microfilms, la televisión, y otras tecnologías incipientes como el fax, las técnicas de reconocimiento de voz, o el procesamiento automatizado de datos con tarjetas perforadas), ayudaría a los científicos a almacenar, seleccionar, recuperar, y compartir información generada en sus investigaciones.

Esta máquina permitiría al investigador almacenar miles y miles de documentos microfilmados a texto completo que podrían ser transmitidos de forma codificada a cualquier otro terminal remoto (que a su vez generaría copias microfilmadas de los documentos, los almacenaría y los proyectaría en la pantalla del usuario).

Además, sería capaz de agregar varios documentos para crear nuevos recursos y anotarlos con información adicional, como notas marginales o comentarios (introduciendo de esta forma la figura del lector-autor que no existía en las bibliotecas tradicionales).

El investigador podría además establecer "*sendas de información*" en las que definiría asociaciones permanentes entre recursos relacionados de su interés, facilitando de esta forma la navegación entre documentos y el posterior acceso a estos patrones de conocimiento. Como vemos, la noción de "*sendas de*

---

*información*” podría considerarse el germen de los enlaces hipertexto, de recursos como las enciclopedias online similares a la Wikipedia (donde la participación activa del usuario-autor es imprescindible), y por extensión, el germen de la propia Web.

No obstante, esta idea de crear un sistema de información universal y ubicuo tiene varios precedentes. En 1941 un profético Jorge Luís Borges presentaba en su cuento *“La biblioteca de Babel”* (Borges 1999) la idea de una biblioteca universal que acumulaba todo el conocimiento del mundo, y como guardiana de esa sabiduría se convertía en una especie de depósito en el que era posible encontrar solución a cualquier tipo de problema.

*“Desde esas premisas indisputables, el bibliotecario dedujo que la Biblioteca es 'total'... es decir, todo lo que se puede expresar, en todas las lenguas. (...) Cuando se anunció que la Biblioteca contenía todos los libros, la primera reacción fue de alegría infinita. Todos los hombres se sintieron poseedores de un tesoro intacto y secreto. No había ningún problema personal, ningún problema mundial, cuya solución elocuente no existiera.”*

Si seguimos retrocediendo en el tiempo, encontramos que más de un lustro antes de la publicación de este relato, Paul Otlet publicaba su *“Tratado de Documentación”* (Otlet 1996), una obra que, desde una perspectiva científica, describía detalladamente conceptos que hoy día nos son muy familiares.

Fruto de la corriente universalista típica de finales del siglo XIX y principios del XX, Otlet propone reorganizar la sociedad mundial de entreguerras creando una red global de comunicación, cooperación e intercambio de información, ideas y conocimiento que denomina *ciudad mundial*.

Esta ciudad sería un ente extraterritorial que se conformaría no solo como un enorme centro de información, educación y enseñanza, sino que al tiempo sería el centro neurálgico mundial de la economía, el comercio y las finanzas.

---

En este contexto, el libro, entendido como el canal primordial de transmisión y difusión del conocimiento humano, ya no sería físicamente igual al tradicional, sino que aglutinaría en sí no solo información textual, sino que otros medios de comunicación, expresión artística e información como el cine, el fonógrafo, la radio, el teléfono, el telégrafo y la televisión abandonarían su papel como sustitutos del propio libro y pasarían a ser sus componentes esenciales.

El libro se convierte así en una extensión de la memoria humana, haciéndola ubicua, universal y eterna, aproximando al hombre, en palabras del propio Otlet, a un estado cercano al de la divinidad.

En su nuevo modelo de sociedad creado entorno a la información Otlet también especula sobre la creación de una máquina que permitiría a la gente acceder a la información disponible en la *"ciudad mundial"*. Esta máquina tendría forma de una mesa de trabajo convencional, pero en lugar de estar ocupada por libros dispondría de una pantalla y un teléfono a través del cual el usuario podría realizar las consultas de información.

La máquina estaría conectada a un edificio remoto en el que un equipo profesionales especializados se encargarían de almacenar, clasificar, distribuir y responder a las consultas de los usuarios. Estos obtendrían como respuesta un conjunto de documentos pertinentes a su necesidad de información que podrían ser consultados de forma simultánea en pantalla, e incluso ser escuchados a través de unos altavoces.

Sin miedo a resultar exagerados, estos autores podrían ser calificados de visionarios que, en su afán por buscar soluciones para satisfacer las ansias de conocimiento del ser humano, trataron de buscar una forma de hacer realidad sus aspiraciones ideando mecanismos y sistemas de información adelantados a la mentalidad y la tecnología de su tiempo, vaticinando de esta manera, con décadas de antelación, la aparición de los ordenadores personales e Internet.

---

Habría que esperar hasta 1969 para que se llevara a cabo la primera conexión entre ordenadores mediante la red ARPANET, a la década de los setenta para que se empezaran a producir en serie los primeros micro-ordenadores, y a 1989 hasta que Tim Berners-Lee desarrollara la World Wide Web.

Poco podían imaginar los investigadores del CERN que la herramienta que estaban utilizando para compartir información científica sería a día de hoy el canal de información de acceso público más dinámico y con mayor tránsito de datos de la historia.

Sin embargo, y a pesar de que la Web ha permitido crear un espacio común de intercambio de información, tal y como predecían Otlet y Bush, de nuevo el gran volumen de recursos que alberga y la ineficiencia de las herramientas de acceso y recuperación de información nos devuelve a la misma cruda situación de desbordamiento que describía Bush en su trabajo hace más de cincuenta años.

Este problema endémico en los sistemas de información se presenta como uno de los principales retos a resolver en la Web debido a sus características específicas.

En la actualidad la información está representada fundamentalmente en lenguaje natural, de forma que los humanos podemos leerla y comprenderla, pero no puede ser directamente interpretada por los ordenadores. Es decir, las máquinas son incapaces de determinar el significado de la información contenida en la Web, recayendo esta tarea en el propio usuario.

Ante esta situación, y en un ejercicio de “clarividencia” similar al de Otlet y Bush, en 1994 Tim Berners-Lee (Berners-Lee 1994) introduce un nuevo elemento como posible solución para paliar los problemas que presenta la Web, aunque difícil de conseguir con las tecnologías existentes en ese momento: la semántica.

---

*“Para un ordenador la Web es un mundo plano y aburrido, desprovisto de significado. Es una pena, ya que de hecho los recursos en la Web describen objetos reales y conceptos imaginarios, y definen relaciones particulares entre ellos. Por ejemplo, un documento puede describir una persona. El título de propiedad de una casa describe, a su vez, una casa y también una relación de propiedad de esa casa hacia una persona. Añadir semántica a la Web implica dos cosas: permitir que los documentos contengan información legible por máquina, y permitir que los enlaces definan relaciones entre recursos. Solo cuando se alcance este nivel extra de semántica podremos usar la capacidad de los ordenadores para sacar provecho de la información más allá de la mera lectura de la misma.” [traducción libre]*

No obstante, y dadas las limitaciones que imponen las técnicas de procesamiento del lenguaje natural (PLN) la solución tampoco pasa por pretender que las máquinas sean capaces de interpretar íntegramente el significado de la información que circula por la red (no sería factible hacer toda la información semánticamente accesible a las máquinas), sino que sería necesario incrustar descripciones semánticas de los recursos en los propios recursos, lo cual los situaría en un marco conceptual determinado que haría posible extraer de ellos conocimiento contextualizado.

Para desarrollar nuevas herramientas y servicios capaces de resolver este problema es necesario que se produzca un salto cualitativo que permita a las máquinas desarrollar una especie de ‘razonamiento automatizado’.

Esto nos lleva a preguntarnos: ¿sería posible representar la información de forma que los ordenadores fueran capaces de interpretarla y ayudarnos de una manera automática a realizar nuestras búsquedas de una manera más precisa?

---

En el trabajo *“The Semantic Web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities”*, Tim Berners-Lee, James Hendler y Ora Lassila (2001) presentan un nuevo escenario, que denominan *Web Semántica*, en el que podría ser posible relajar las estrictas restricciones que imponen las soluciones *ad-hoc* que presentan los sistemas de representación del conocimiento y adaptarlas a la propia idiosincrasia de la Web.

La gran diferencia que existe entre este modelo y las propuestas de Otlet y Bush radica en que, mientras estos se limitaron a plantear un escenario hipotético de desarrollo apoyados sobre las tecnologías existentes en la época, Berners Lee establece unas bases sólidas para el desarrollo de nuevas herramientas tecnológicas y metodológicas para la realización del proyecto.

No obstante, la forma en que evolucionará el modelo, y los vocabularios que acabarán triunfando entre usuarios y desarrolladores es aún una incógnita.

A continuación vamos a ofrecer una idea general sobre la estructura, principales características, tecnologías y objetivos que persigue el modelo de Web Semántica.

---

## 2.2 ¿QUÉ ES LA WEB SEMÁNTICA?

La Web de hoy día permite a la gente el acceso a recursos y servicios en Internet. No obstante, muchos de los procesos necesarios para acceder a estos recursos y servicios requieren de la intervención intelectual consciente y crítica del usuario (como por ejemplo, decidir qué recursos son de su interés de entre un listado de documentos recuperados por un buscador). Los interfaces de los servicios están descritos en lenguaje natural y antes de interactuar con ellos necesitan que el usuario realice un ejercicio de comprensión para poder interactuar con ellos.

La Web Semántica supone una extensión de la Web actual, donde la información está dotada de un significado bien definido y puede ser procesada por las máquinas e interpretada por los humanos, permitiendo de esta forma una mejor cooperación entre ellos (Berners-Lee, Hendler, Lassila 2001).

Es un dominio transversal que se apoya sobre múltiples áreas de conocimiento de las que se sirve, y que al mismo tiempo contribuyen activamente a su desarrollo e implantación. Entre ellas encontramos las siguientes (Euzenat 2002):

- *Inteligencia artificial:* Aporta, por ejemplo, mecanismos de razonamiento, lenguajes de representación del conocimiento, procesamiento de información aproximado y difuso, o técnicas de descubrimiento de recursos.
  - *Desarrollo Web:* Utiliza mecanismos de modelado de usuarios y de identificación, y tecnologías y lenguajes basados en XML que son básicos en la Web Semántica.
  - *Bases de datos:* La Web Semántica adopta de las bases de datos los mecanismos básicos para el almacenamiento eficiente de los datos, para el control de la tolerancia a fallos o para gestionar la seguridad. Además, algunos lenguajes de consulta como SQL han servido de modelo para el
-

desarrollo de muchos de los lenguajes de interrogación semánticos que existen hoy día.

- *Agentes software*: De este campo la Web Semántica adopta el procesamiento distribuido, los lenguajes de comunicación y la aplicación de protocolos de interacción y cooperación.
  - *Informática teórica y computacional*: La aportación de este campo se traduce en el uso de lenguajes de programación y plataformas de desarrollo de aplicaciones basadas en estos lenguajes (como, por ejemplo, la plataforma Jena [13], en la utilización de mecanismos de deducción automática (como la lógica descriptiva que utilizan gran parte de los lenguajes ontológicos de la Web Semántica) y en la definición de semánticas.
  - *Ingeniería de sistemas*: Desarrolla mecanismos para garantizar la fiabilidad y la seguridad en la Web, así como para la utilización de plataformas móviles.
  - *Lingüística computacional y reconocimiento de patrones*: Son herramientas útiles para la Web Semántica ya que definen las bases de la adquisición automatizada de conocimiento a partir de fuentes primarias, y del uso de recursos léxicos para el desarrollo de ontologías. Además son de utilidad, entre otras cosas, para el diseño de sistemas de pregunta-respuesta y para la definición de normas pragmáticas.
  - *Ingeniería de documentos y bibliotecas digitales*: Definen los patrones básicos para la transformación, etiquetado e indización de recursos.
  - *Desarrollo de interfaces*: De esta área la Web Semántica aprovecha, por ejemplo, los principios del trabajo colaborativo asistido por ordenador, la evaluación de factores de trabajo y los estudios de comunicación.
-



- *Ciencias sociales y humanidades*: Sirven de soporte principalmente en procesos de validación de ontologías.

Apoyado sobre estos pilares, el fin último del modelo es que las máquinas puedan asistir de una manera más eficiente a los humanos en sus tareas de acceso a recursos y servicios, y la creación de un sistema de información más abierto que permita el desarrollo de nuevos servicios y aplicaciones.

Se basa en dos ideas principales: el etiquetado semántico de recursos (lo que implica una separación formal entre el contenido y la estructura de los documentos), y la creación de aplicaciones software “inteligentes” (también conocidas como *agentes*) capaces de procesar y operar con estos recursos a nivel semántico (Hendler 2001).

Los agentes son entidades software que realizan procesos de recogida, filtrado, procesamiento de información e inferencia de una forma semiautónoma. Es decir, el agente software es capaz de interpretar adecuadamente lo que se le pide, de comprender el contenido de los recursos, de validar y evaluar la relevancia de los resultados que obtiene y de obtener nuevo conocimiento. Cumplen el papel de *infomediarios*, es decir, intermediarios entre las necesidades de los usuarios y las fuentes de información distribuidas disponibles en la red. De hecho, el verdadero potencial de la Web Semántica se desarrollará cuando la web esté *poblada* por un gran número de agentes que sean capaces de recoger información de diferentes fuentes distribuidas, procesarla e intercambiar resultados con otros agentes.

No obstante, para que los agentes puedan desarrollar sus tareas en la red de una manera eficiente es necesario que la información con la que trabajan sea interoperable, es decir, debe estar representada (tanto a nivel sintáctico como semántico) de forma que pueda ser reutilizada tanto como sea posible. Basado en este principio, Berners-Lee (2000) desarrolló un modelo multicapa (ver Fig. 4), en el que se procura la máxima interoperabilidad de cada capa con su inmediatamente inferior y su inmediatamente superior (esto implica que los

---

recursos definidos en una capa determinada pueden ser reutilizados íntegramente por las capas superiores, aunque sólo parcialmente por las inferiores).

El modelo está formado por seis capas diferentes: en las tres primeras se establecen las bases para poder representar la información de una manera *semánticamente* accesible, mientras que en las tres capas superiores se definen los elementos que permiten a los agentes software autentificar y comprobar la confiabilidad de los diferentes elementos del modelo (recursos, agentes, inferencias obtenidas, etc.).

A continuación pasamos a describir cada una de estas capas y los elementos básicos que en ellas se definen.

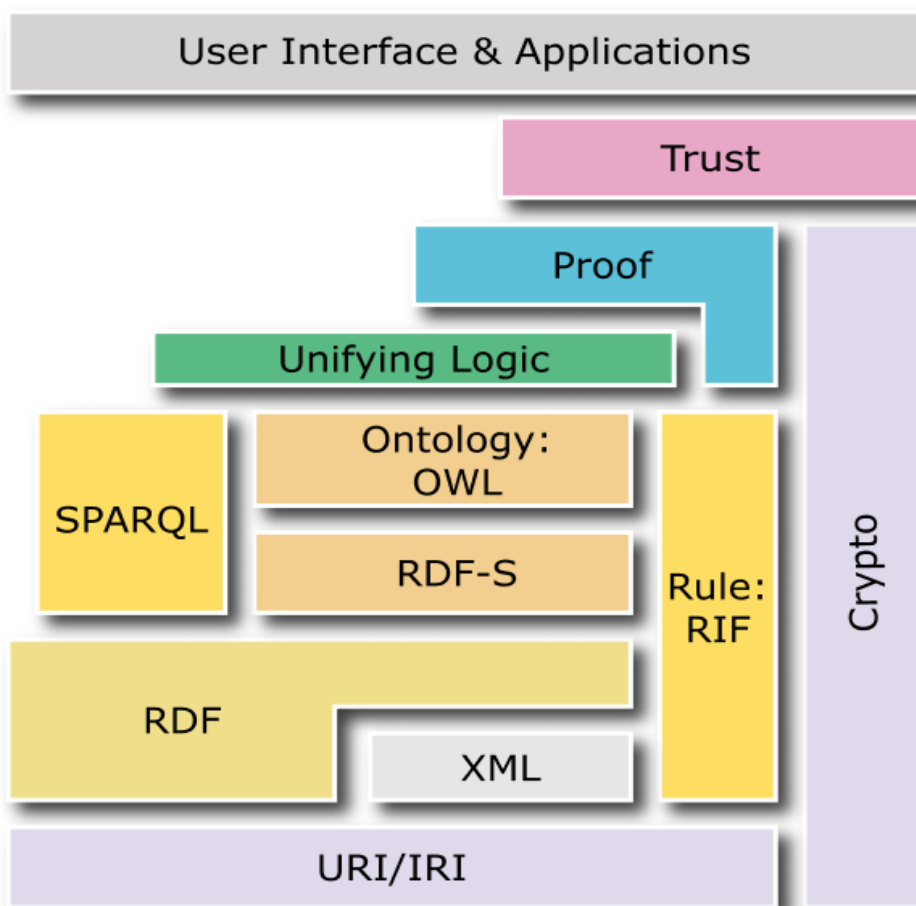


Fig.4 Niveles de la Web Semántica

[extraído de <http://www.w3.org/2007/Talks/0130-sb-W3CTechSemWeb/layerCake-4.png>]

### 2.2.1 CAPA SINTÁCTICA

Esta capa supone el basamento del modelo de Web Semántica y en ella se definen una serie de elementos que permiten el intercambio y reutilización de recursos de fuentes de información heterogéneas y distribuidas. Así, se establece el estándar Unicode [14] como patrón de codificación de caracteres universal, y las URI (*Uniform Resource Identifier*) (Berners-Lee, Fielding, Masinter 2005) como esquema para identificar recursos de una forma unívoca. Esta identificación se puede realizar tanto a través de URLs (*Uniform Resource Locator*) que describen la localización física de un recurso determinado, como de URNs (*Uniform Resource Name*) que identifican el recurso unívocamente, independientemente de su ubicación física. Las URI's aprovechan la convención para definir direcciones o rutas sobre una estructura de directorios en la web, independientemente de que exista o no un documento físicamente asociado a esa URL.

También es necesario definir un formato estándar válido para representar la información. Para ello recurrimos al lenguaje de marcado XML (*eXtensible Markup Language*) (Bray et al. 2006), que va a conformar la base sintáctica de todo el modelo. Este metalenguaje es una adaptación simplificada de SGML (*Standard Generalized Markup Language*) que permite de una forma simple y flexible la descripción normalizada de recursos mediante conjuntos de etiquetas, lo que lo convierte en un sistema adecuado para definir, validar y compartir recursos en la red.

Por ejemplo, si quisiéramos crear un catálogo de los libros que tenemos en casa podríamos crear un documento XML como el de la Fig.5. En él se define el elemento '*CatálogoLibros*', y dentro de este un sub-elemento denominado '*libro*' que a su vez está caracterizado por una serie de elementos y atributos (*isbn*, *título*, *autor*, *materia*, *resumen*, etc.), que toman un valor determinado (una cadena de caracteres o literal).

---

Como se puede observar, es posible definir con total libertad el conjunto de etiquetas más adecuado para resolver un problema determinado, en este caso, describir el contenido de los elementos de un catálogo. Este conjunto de etiquetas necesita un nombre que las identifique unívocamente para distinguirlas de cualquier otro conjunto de etiquetas definido por otra persona. Para ello utilizamos los espacios de nombre (Bray, Hollander, Layman 1999), un sencillo método de identificación de conjuntos de etiquetas mediante URIs. Cada vocabulario estará identificado por una URI y un prefijo que identifica a todas las etiquetas pertenecientes a un mismo conjunto de etiquetas. De esta forma es posible, por ejemplo, utilizar simultáneamente en un mismo documento varios vocabularios sin que exista riesgo de colisión entre ellos.

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<CatalogoLibros xmlns="http://perso.wanadoo.es/kishimaru/Ejemplo">
  <libro isbn="_84-473-0224-5">
    <titulo>Los dragones del Edén</titulo>
    <autores>
      <autor>Carl Sagan</autor>
      <traductor>Jaume Josa Llorca</traductor>
    </autores>
    <editorial>RBA</editorial>
    <materia>Divulgación</materia>
    <resumen>Relato de los hechos fundamentales en el desarrollo de la vida
      inteligente sobre la Tierra, en el que se analiza desde la cantidad
      de información contenida en un cromosoma, hasta posibles líneas de
      evolución del cerebro humano.
    </resumen>
  </libro>
</CatalogoLibros>
```

Fig.5 Ejemplo de documento XML

Una vez que tenemos representada la información contenida en nuestros documentos XML, sería muy útil poder definir y validar tanto su contenido como su estructura. De esta forma podemos establecer las reglas sintácticas para crear un vocabulario que solucione un problema determinado (como por ejemplo, el conjunto de etiquetas válidas o el orden en el que estas deben aparecer en el documento), y la estructura válida de todos los documentos que se ajusten a dicho vocabulario (es decir, podemos crear una tipología documental

determinada). Para desarrollar esta tarea podemos utilizar diferentes lenguajes de esquema sintácticos como las DTD (*Document Type Definiton*) y XML Schema.

Una DTD, tal y como su nombre indica, es la definición de un tipo de documento, es decir, la definición de una serie de requisitos para que nuestro documento XML sea considerado válido. Con una DTD podemos definir, por ejemplo, la validez de una etiqueta, su obligatoriedad, su ocurrencia e incluso el valor que puede tomar. Ejemplos de DTD serían HTML, que no es más que un tipo de documento de SGML, o EAD (*Encoded Archival Description*) [15] un vocabulario XML específico para descripción de archivos.

El diseño de una DTD es realmente sencillo ya que consiste en definir una serie de reglas y restricciones (utilizando una sintaxis no XML) sobre los diferentes elementos, atributos y valores que se pueden declarar en un documento.

```
<!ELEMENT CatalogoLibros (libro+)>
<!ELEMENT libro (titulo, autores, editorial, materia, resumen?)>
<!ATTLIST isbn CDATA #REQUIRED>
<!ELEMENT titulo (#PCDATA)>
<!ELEMENT autores (autor+, traductor)>
<!ELEMENT autor (#PCDATA)>
<!ELEMENT traductor (#PCDATA)>
<!ELEMENT editorial (#PCDATA)>
<!ELEMENT materia (Divulgación | Novela historica | Ciencia-Ficción)>
<!ELEMENT resumen (#PCDATA)>
```

Fig.6 Ejemplo de DTD

No obstante, las DTDs no son adecuadas para resolver problemas algo más complejos, como por ejemplo especificar el formato que debe tener un tipo de dato específico (como la estructura de un ISBN o una dirección de correo electrónico), ya que las DTDs solo soportan un número limitado de tipos de datos básicos. Por ello, para estructurar y validar documentos más complejos (o para hacerlo de una forma más precisa) es necesario recurrir a XML Schema (Fallside, Walmsley 2004), una herramienta más potente que las DTD, con capacidad para crear nuevos tipos de datos (extendiendo, restringiendo o reutilizando otros tipos de datos nativos de la especificación), y para controlar la recurrencia de los

---

elementos y atributos de los documentos. Además cuenta con la ventaja añadida de utilizar sintaxis XML (lo que evita tener que manejar varias sintaxis diferentes y puede facilitar la interpretación de su código).

En el siguiente ejemplo vemos cómo se definiría el tipo de dato ISBN para los libros publicados en España (ejemplo extraído de Costello y Sperberg (2001)).

```
<xsd:pattern value="84-\d{0-9}|-}{6}\d-[0-9x]">
  <xsd:annotation>
    <xsd:documentation>
      grupo/país ID = 84 (guión despues del segundo dígito)
      País = España. Dígito de control 0-9 o 'x'
    </xsd:documentation>
  </xsd:annotation>
</xsd:pattern>
<xsd:pattern value="84\s\d{0-9}|\s}{6}\d\s[0-9x]">
  <xsd:annotation>
    <xsd:documentation>
      grupo/país ID = 84 (espacio después del segundo dígito)
      País = España Dígito de control 0-9 o 'x'
    </xsd:documentation>
  </xsd:annotation>
</xsd:pattern>
```

Fig.7 Ejemplo de definición de un tipo de dato con XML Schema

### 2.2.2 CAPA SEMÁNTICA

Además de la flexibilidad, independencia, consistencia, capacidad descriptiva y discriminatoria que proporcionan las herramientas definidas en la capa sintáctica, el modelo necesita de un medio para dotar de semántica a los recursos Web. Para definir un modelo semántico es necesario identificar los recursos de forma unívoca, caracterizar los metadatos de los documentos, y crear un modelo lógico de metadatos a partir del cual se pueda inferir conocimiento. Con este fin, el W3C ha desarrollado las recomendaciones RDF (*Resource Description Framework*) y RDFS (*RDF Schema*).

RDF (Beckett 2004) es un lenguaje que permite codificar, intercambiar y reutilizar metadatos estructurados. Se puede considerar como la piedra angular sobre la que se vertebra la estructura semántica de la Web. Este lenguaje añade

semántica a la información estructurándola en forma de triplas sujeto-propiedad-objeto (donde el valor puede ser otro sujeto o un literal) que pueden ser representadas en forma de grafos (de hecho es habitual referirse a los documentos RDF como grafos (ver Fig.8 ).

La ventaja de utilizar RDF frente a lenguajes de etiquetado *no semánticos*, como XML, es que mientras estos se limitan a definir un vocabulario que permite describir la información para que sea directamente interpretable por humanos, RDF da la vuelta de tuerca necesaria para que esa información sea también interpretable y procesada por los agentes software.

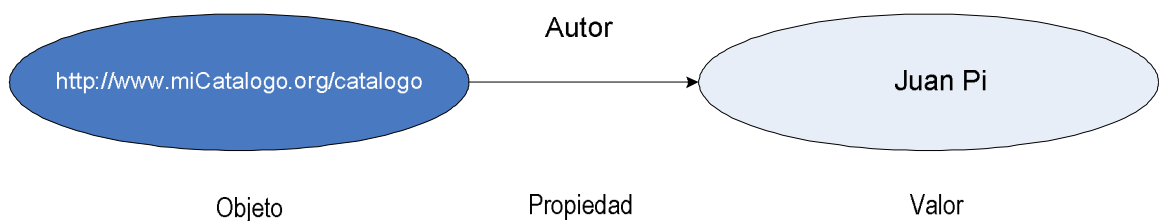


Fig.8 Ejemplo de grafo RDF

RDF puede ser expresado utilizando diferentes sintaxis, como Notation3 (Berners-Lee 2006) o XML, pero no cabe duda de que la serialización XML facilita la migración entre estos dos formatos, ya que solo es necesario adaptar los documentos XML preexistentes a una nueva estructuración lógica. En la Fig.9 vemos una posible forma (no la única) de expresar el catálogo de libros de la Fig.8 en formato RDF:

---

```

<Catalogo rdf:ID="CatalogoLibros"
xsi:schemaLocation="http://perso.wanadoo.es/kishimaru/Ejemplo_CataSchemaRDF.xsd">
  <registro>
    <Libro rdf:ID="_84-473-0224-5">
      <titulo>Los dragones del Edén</titulo>
      <autores>
        <rdf:Description>
          <autor>Carl Sagan</autor>
          <traductor>Jaume Josa Llorca</traductor>
        </rdf:Description>
      </autores>
      <editorial>RBA</editorial>
      <materia>Divulgación</materia>
      <resumen>Relato de los hechos fundamentales en el desarrollo de la vida
        inteligente sobre la Tierra, en el que se analiza desde la
        cantidad de información contenida en un cromosoma, hasta posibles
        líneas de evolución del cerebro humano.
      </resumen>
    </Libro>
  </registro>
</Catalogo>

```

Fig.9 Ejemplo de documento RDF

En este ejemplo concreto, dentro del recurso *'Catálogo'* se anida la propiedad *'registro'* cuyo valor es el recurso *'Libro'*. Este a su vez contiene las propiedades *'título'*, *'resumen'*, etc., cada una con su valor correspondiente (ya sea un literal u otro recurso). Y seguiríamos anidando así recursos, atributos y valores sucesivamente según la estructura y complejidad lógica de la información que estemos representando.

RDF Schema (Brickley, Guha 2002) es una extensión semántica de RDF, y define un lenguaje con el que es posible construir taxonomías de dominios particulares, definiendo clases y propiedades, relaciones básicas entre ellas (básicamente jerárquicas), y restricciones de rango y dominio para esas propiedades. Sin embargo, y a pesar de lo que pudiera parecer, RDFS no es un lenguaje lo suficientemente potente como para poder definir ontologías (van Ossenbruggen, Hardman, Rutledge 2002). Por lo tanto, RDFS se podría definir como un lenguaje de esquema *"semántico"* (para diferenciarlo de un lenguaje de esquema *"sintáctico"* como XML Schema) cuyo papel consiste en ofrecer información adicional sobre la interpretación que se debe hacer de los asertos definidos en los documentos RDF.



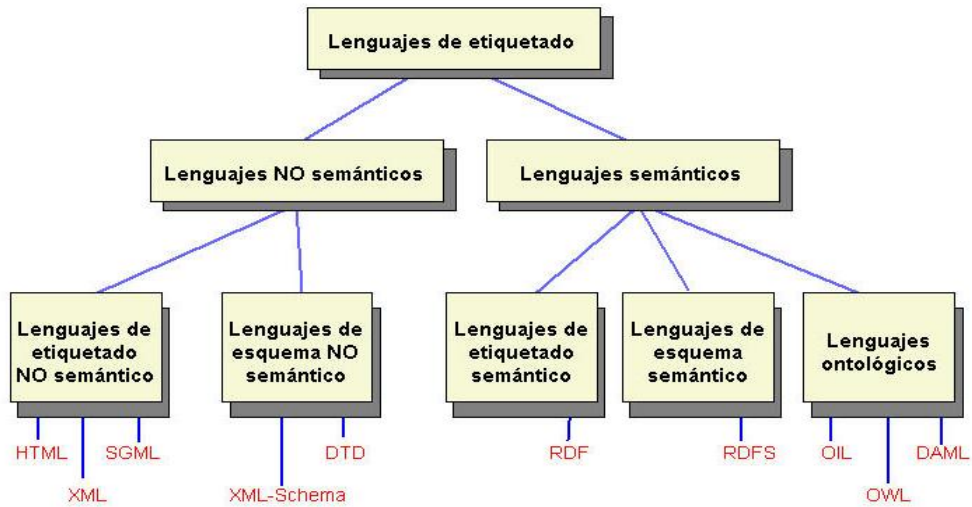


Fig.10 Ejemplo de clasificación jerárquica

Veamos un ejemplo para explicar su función dentro del modelo. Supongamos un sitio web académico cuya línea de investigación se centra en estudiar los lenguajes de etiquetado orientados a la Web. Una posible taxonomía de este dominio del conocimiento podría ser la que representamos en el árbol jerárquico de la Fig.10.

```

<rdfs:Class name="Lenguajes_Etiquetado"/>
<rdfs:Class name="Lenguajes_Semanticos">
<rdfs:subClassOf rdf:resource="#Lenguajes_Etiquetado"/>
</rdfs:Class>
<rdfs:Class name="Etiquetado_Semantico">
<rdfs:subClassOf rdf:resource="#Lenguajes_Semanticos"/>
</rdfs:Class>
<rdfs:Class name="Lenguajes_Eschema_Semantico">
<rdfs:subClassOf rdf:resource="#Lenguajes_Semanticos"/>
</rdfs:Class>
<rdfs:Class name="Lenguajes_Ontologicos">
<rdfs:subClassOf rdf:resource="#Lenguajes_Semanticos"/>
</rdfs:Class>
...

<rdf:Property rdf:ID="Define_Ontologias_Modelado">
<rdfs:domain rdf:resource="#Lenguajes_Ontologicos"/>
<rdfs:range rdf:resource="#Etiquetado_Semantico"/>
</rdf:Property>
...

```

Fig.11 Ejemplo de documento RDFS

En la Fig.11 presentamos un ejemplo de esquema RDF aplicado a esta taxonomía (por simplicidad, solo mostramos cómo expresar las entidades y dependencias jerárquicas que establecen entre si las clases de la rama derecha del árbol).

Pero no solo basta con definir una estructura para dotar la información de semántica, sino que es necesario poder realizar búsquedas y extraer información de los grafos RDF. Para ello se han desarrollado diferentes lenguajes de consulta e inferencia cuya principal diferencia con los lenguajes de búsqueda tradicionales radica en su capacidad para equiparar tanto literales como estructuras semánticas (Shah et al. 2002) (Guha, McCool, Miller 2003).

Entre ellos podemos encontrar SerQL (*Sesame RDF Query Language*) (Aduna, Sirma 2005), TRIPLE [16], RQL (*RDF Query Language*) [17], eRQL (*easy RDF Query Language*) [18], Xcerpt [19] que es un lenguaje basado en reglas, o RDQL (*RDF Data Query Language*) (Seaborne 2004).

No obstante, la propuesta del W3C (de hecho recomendación desde este año 2008) es SPARQL (*SPARQL Protocol and RDF Query Language*) un lenguaje semántico de consulta (Prud'hommeaux, Seaborne 2008) y protocolo de acceso y recuperación de datos (Clark, Feigenbaun, Torres 2008). Tiene una estructura similar a SQL, y dispone de herramientas capaces de gestionar diferentes tipos de datos de múltiples fuentes distribuidas (independientemente de si los datos están definidos en origen en formato RDF o están representados en este formato a través de aplicaciones intermediarias) y realizar búsquedas directas en grafos RDF cuyos resultados pueden ser un conjunto de datos (expresados por ejemplo en formato XML o HTML) o nuevos grafos RDF. SPARQL tiene además capacidad para realizar consultas en patrones alternativos de grafos (como las conjunciones o disyunciones de grafos) así como para restringir los valores que pueden aparecer como resultado, o definir búsquedas restringidas a grafos de recursos determinados.

---

Un ejemplo sencillo de la sintaxis de este lenguaje es el siguiente:

```
PREFIX ex: <http://example.org/ex#>
SELECT ?titulo
WHERE ?x ex:titulo ?titulo
```

Fig.12 Ejemplo de consulta definida con SPARQL

En este caso, el sistema recuperaría el valor (*?titulo*) de la propiedad *ex:titulo* de todos los sujetos que tengan esa propiedad (genéricamente representados por la variable *?x*).

### 2.2.3 CAPA ONTOLÓGICA

Una vez que la información la tenemos estructurada de forma que pueda ser procesada por máquina, es necesario contextualizarla dentro de un dominio concreto para poder inferir conocimiento a partir de ella.

Es en la capa ontológica donde podemos definir ontologías, entendidas como la suma de una serie de conceptos relevantes del conocimiento compartido por los miembros de un dominio concreto, las relaciones que establecen entre sí estos conceptos, y los axiomas definidos sobre estos conceptos y relaciones.

De una manera más simple, podríamos definir “*ontología*” como la conceptualización formal de una parcela de realidad o un dominio con la que pueden operar diferentes aplicaciones software. De esta manera, los conceptos o términos utilizados para la descripción sirven como un vocabulario común (sintáctico y semántico) que favorece la comunicación y la interoperabilidad de recursos entre aplicaciones. Dan sentido pleno a la información situándola en un contexto, permitiéndonos dar un salto cualitativo muy importante: de utilizar motores de búsqueda que trabajan con palabras clave, pasaríamos a utilizar agentes software inteligentes que trabajan con conceptos. Esto supondría, a su

---

vez, pasar de la mera recuperación de información a la obtención de respuestas precisas a consultas concretas, es decir, a recuperar conocimiento.

En la actualidad existen diferentes lenguajes formales para la definición e instanciación de ontologías en la Web, como DAML (*DARPA Agent Markup Language*) [20], OIL (*Ontology Inference Layer*) (Fensel 2001), DAML+OIL (Connolly et al. 2001), y OWL (*Web Ontology Language*) (McGuinness, Harmelen 2004), que es la recomendación del World Wide Web Consortium (W3C). OWL se define como una extensión semántica de RDFS, que permite definir clases y propiedades complejas para diseñar ontologías web manteniendo la convención sintáctica de RDF.

En OWL se distinguen tres versiones o “especies” (van Harmelen, McGuinness 2004) que se diferencian por su capacidad expresiva: OWL Lite es la versión más simple y en ella se definen una serie de elementos básicos para crear de una manera simple ontologías fáciles de procesar; OWL DL incluye todas las funcionalidades del lenguaje pero imponiendo una serie de restricciones sobre las propiedades de RDF que pueden ser usadas; la versión más completa es OWL Full, que incluye todas las funcionalidades del lenguaje y admite propiedades típicas de RDF como la reificación, que permite definir sentencias sobre sentencias.

A la hora de desarrollar cualquier aplicación semántica es importante tener en cuenta las características de cada versión para elegir aquella que mejor se ajusta a nuestras necesidades específicas, ya que cuanto mayor es la capacidad expresiva de una versión, mayor será también la dificultad que tendrán los agentes para operar con ellas de una manera eficiente.

Dado el nivel de desarrollo actual del modelo, para garantizar la operatividad de las aplicaciones se recomienda utilizar OWL Lite u OWL DL.

---

```
<owl:Class rdf:ID="Sintaxis_NO_XML"/>
<owl:Class rdf:ID="Sintaxis_XML">
  <owl:intersectionOf rdf:parseType="Collection">
    <owl:Class rdf:about="Lenguajes_Etiquetado"/>
    <owl:Class>
      <owl:complementOf rdf:resource="#Sintaxis_NO_XML"/>
    </owl:Class>
  </owl:intersectionOf>
</owl:Class>
```

Fig.13 Ejemplo de clase descrita en OWL

Para ejemplificar la capacidad de OWL para componer clases complejas, en la Fig. 13 definimos la clase '*Sintaxis XML*' como la intersección de la clase '*Lenguajes de etiquetado*' y el complemento lógico de la clase '*Sintaxis NO XML*'. O dicho de otra forma, estamos considerando dentro de la clase '*Sintaxis XML*' a todos aquellos lenguajes de etiquetado que no pertenecen a la clase '*Sintaxis NO XML*'.

#### 2.2.4 CAPA LÓGICA

Sobre la capa ontológica se apoyan las tres últimas capas del modelo, en las que se tratan básicamente cuestiones relacionadas con la seguridad, veracidad y confianza de la información que intercambian entre sí los agentes software. Dado que el modelo de Web Semántica está a penas en su fase inicial, el nivel de desarrollo de estas tres capas es escaso aún, pero no obstante vamos a repasar los elementos fundamentales que se definen en cada una de ellas.

La primera de estas capas es la capa lógica, y en ella es donde se establecen las diferentes reglas de producción con las que describir el conjunto de deducciones que se pueden hacer a partir de un conjunto determinado de datos. En otras palabras, en esta capa se establecen los pasos que un agente debe seguir para llegar a una conclusión con las inferencias obtenidas en un proceso de búsqueda de información.

---

Estas reglas adoptan la forma de implicaciones entre un antecedente y un consecuente, y se pueden definir usando diversos lenguajes como RuleML (*Rule Markup Language*) [21], o SWRL (*Semantic Web Rule Language*) (Horrocks 2003) (que permite integrar las reglas dentro del mismo código de las ontologías definidas en OWL Lite o DL).

Veamos un ejemplo. Supongamos que queremos definir una regla en un sistema de información, que defina el procedimiento que debe seguir el agente del sistema cuando un usuario solicite un libro cuya materia sea “*Divulgación*”. Cuando se de esta situación, el agente del sistema deberá recomendar un libro (distinto del que ha escogido el usuario) que pertenezca a la categoría “*Novedades de divulgación*”. De una forma informal esta regla se podría expresar de la siguiente manera:

$$[Solicita (usuario, libro1)] \wedge [Materia (libro1, Divulgación)] \wedge [libro2 \in Novedades\_Divulgación] \wedge [libro1 \neq libro2] \Rightarrow Recomendar (usuario1, libro2)$$

Como vemos, hay un antecedente donde se definen una serie de requisitos que es necesario que sean ciertos (es decir, que se satisfagan) para que el consecuente también sea cierto y por lo tanto la regla se pueda aplicar. Si la traducimos a SWRL, la regla se expresaría tal y como aparece en la Fig.14.

Como resultado el agente personal del usuario recibiría la referencia del libro que estaba buscando y además otra referencia de un libro recomendado por el agente del sistema.

Un área en la que también se está mostrando un gran interés por el desarrollo de vocabularios que permiten definir reglas es la de los servicios web. Los servicios web son sitios web que no solo presentan información estática, sino que permiten al usuario realizar algún tipo de acción como comprar un producto o controlar un artefacto de forma remota (Alonso 2004). Por esta razón es de vital importancia para los proveedores de este tipo de servicios describirlos de una manera exhaustiva para facilitar los procesos de localización, invocación, uso, seguimiento e interacción con otros servicios web (McIlraith, Son, Zeng 2001).

```

<swrl:Variable rdf:ID="usuario"/>
<swrl:Variable rdf:ID="libro1"/>
<swrl:Variable rdf:ID="libro2"/>
<ruleml:imp>
  <ruleml:_body rdf:parseType="Collection">
    <swrl:individualPropertyAtom>
      <swrl:propertyPredicate rdf:resource="solicita"/>
      <swrl:argument1 rdf:resource="#usuario"/>
      <swrl:argument2 rdf:resource="#libro1"/>
    </swrl:individualPropertyAtom>
    <swrl:DatavaluedPropertyAtom>
      <swrl:propertyPredicate rdf:resource="materia"/>
      <swrl:argument1 rdf:resource="#libro1"/>
      <swrl:argument2 rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
        Divulgación
      </swrl:argument2>
    </swrl:DatavaluedPropertyAtom>
    <swrl:ClassAtom>
      <swrl:classPredicate rdf:resource="Novedades_Divulgacion"/>
      <swrl:argument1 rdf:resource="#libro2"/>
    </swrl:ClassAtom>
    <swrlb:notequal>
      <swrl:argument1 rdf:resource="#libro1"/>
      <swrl:argument2 rdf:resource="#libro2"/>
    </swrlb:notequal>
  </ruleml:_body>
  <ruleml:_head rdf:parseType="Collection">
    <swrl:individualPropertyAtom>
      <swrl:propertyPredicate rdf:resource="recomendar"/>
      <swrl:argument1 rdf:resource="#usuario"/>
      <swrl:argument2 rdf:resource="#libro2"/>
    </swrl:individualPropertyAtom>
  </ruleml:_head>
</ruleml:imp>

```

Fig.14 Ejemplo de regla descrita en SWRL

Este tipo de descripciones solo es posible desarrollarlas dentro de un marco como el que proporciona la Web Semántica ya que, a diferencia de otros lenguajes de descripción de servicios web basados en XML, como WSDL (*Web Services Description Language*) (Christensen et al. 2001), los lenguajes semánticos de descripción de servicios web se basan en la utilización de ontologías donde se describen las características y los procesos que se llevan a cabo en el servicio, y en las que es posible incluir reglas.

Entre estos lenguajes encontramos OWL-S (Martin et al. 2004), Semantic Web Services Language (SWSL) (Battle et al. 2005), o Web Service Modeling Language (WSML) (de Bruijn et al. 2005).

### 2.2.5 CAPAS DE PRUEBA Y DE CONFIANZA

Uno de los elementos clave de la Web Semántica es el establecimiento de mecanismos de seguridad adecuados que permitan autenticar y comprobar la confiabilidad de los diferentes elementos del modelo (desde los recursos, a los agentes, las entidades emisoras o los propios individuos) para de esta forma garantizar la fiabilidad de los resultados obtenidos en un proceso de búsqueda de información. Para desarrollar estos mecanismos se definen dos nuevas capas en el modelo: la capa de prueba y la capa de confianza.

No obstante, los modelos y lenguajes que permiten el desarrollo de las capas de prueba y confianza son los que aún se encuentran en un estado más embrionario.

La capa de prueba es la inmediatamente superior a la capa lógica y permite definir una infraestructura adecuada para que los agentes puedan establecer relaciones lógicas complejas con otros agentes, intercambiando pruebas (o explicaciones) sobre las inferencias obtenidas en una búsqueda (Finin, Joshi 2002).

Esta infraestructura se apoya básicamente en tres elementos: i) las reglas de inferencia definidas en la capa lógica donde se especifican los requerimientos de seguridad; ii) la capacidad de los agentes de rastrear y probar el origen de una secuencia lógica gracias a la estructura en que está representada la información (Antinou, van Harmelen 2004A); y iii) las firmas digitales.

Las firmas digitales son bloques de datos cifrados que permiten a los agentes verificar que determinada información proviene de una fuente fiable (Berners-Lee, Hendler, Lassila 2001). A todo documento que va a ser firmado digitalmente se le aplica un algoritmo de “*hashing*” que devuelve un bloque de datos que representa a dicho documento. Este bloque es posteriormente encriptado mediante un sistema de Clave Pública (Kant, Iyer, Mohapatra 2000) que permite

---



identificar al autor/emisor del documento y comprobar la integridad del contenido (es decir, que el documento no ha sido alterado fraudulentamente).

Sin embargo, el uso de firmas digitales presenta dos inconvenientes: por un lado es necesario validar la encriptación de Clave Pública a través de un certificado emitido por una autoridad de certificación (centralizada), lo cual choca directamente con la naturaleza descentralizada de la Web Semántica.

Por otro lado, las firmas digitales no permiten establecer la confianza que a un individuo le merece determinada fuente o agente (y consecuentemente el contenido de los recursos e inferencias que se pueden obtener de ellos).

La solución a estas deficiencias pasa por la definición de una última capa, la capa de confianza, en la que se establecen extensas redes sociales de confianza (Matthew, Agrawal, Domingos 2003) que permitirán a los agentes determinar la confiabilidad de una determinada fuente o recurso y comprobar la autoría de los asertos RDF.

En estas redes todo usuario de la Web está identificado por su propia URI y expresa el grado de confianza o desconfianza (Guha et al. 2004) que le merecen aquellos individuos que conoce (o aquellos con los que a tenido algún tipo de contacto).

Estos grados de confianza podrán definirse, por ejemplo, utilizando una extensión de FOAF (*Friend Of A Friend*) (Brickley, Millar 2005), un vocabulario RDF específico para describir relaciones sociales en la Web (Hendler, Goldbeck, Parsia 2002).

De esta manera, aplicando conjuntamente las reglas de inferencia definidas en la capa lógica, las firmas digitales y las redes de confianza, sería posible construir la denominada "*Web of Trust*" (Web de confianza) en la que se dispondría de mecanismos de autenticación y validación precisos. Así, ya no sería necesario recurrir a la encriptación de datos y, consecuentemente, se podría prescindir de las autoridades de certificación, resolviendo así el problema de la centralización (Reagle 2002).

---

Un ejemplo de aplicación a pequeña escala de la Web of Trust son los sistemas de recomendación semánticos basados en confianza (Ziegler 2004) en los que una serie de agentes generan recomendaciones a sus usuarios de acuerdo al nivel de confianza que les merecen el resto de usuarios.

### **2.2.6 TODA LA MAQUINARIA EN FUNCIONAMIENTO**

Situémonos en el más optimista de los escenarios, un escenario en el que en Internet todos los recursos están descritos en formato RDF y todos los dominios de conocimiento disponen de su correspondiente ontología web.

Supongamos que somos los gestores de una web académica especializada en lenguajes de etiquetado orientados a la Web. Nuestros recursos están accesibles en la red codificados en forma de documentos RDF y de acuerdo con la ontología que han desarrollado reputados organismos que son una autoridad en el tema. Como sólo nos interesan los artículos científicos y las contribuciones a congreso sobre el tema, controlamos la estructura de nuestros recursos utilizando para ello varios XML Schemas que validan y definen ambos tipos de documentos.

En otro lugar del mundo, un estudiante de documentación llamado Marco necesita información sobre las distintas especies de OWL, para realizar un trabajo de clase. Accede al interfaz de búsqueda de su agente software personal y le lanza la consulta, estableciendo unos criterios mínimos de fiabilidad y relevancia de los documentos a recuperar.

El agente tiene acceso a una ontología que controla el perfil de usuario de Marco y comprueba que debe realizar la búsqueda dentro del dominio de los lenguajes de etiquetado, descartando otros dominios. De esta forma se asegura de no recuperar, por ejemplo, recursos sobre ornitología (recordemos que, en inglés, *owl* significa búho).

El agente de Marco se encarga de interactuar con otros agentes de confianza (entre ellos el nuestro) a los que traslada su consulta. El agente de nuestra web

---

evalúa los criterios de búsqueda y comprueba que existen una serie de documentos o partes de documentos que son relevantes para la consulta. Esto se lo hace saber al agente de Marco, el cual pide a nuestro agente pruebas del razonamiento seguido para comprobar que le estamos ofreciendo una información pertinente.

Una vez hecha esta comprobación, con todos y cada uno de los agentes consultados, el agente de Marco recopila e integra la información obtenida, presentando en pantalla el resultado de la búsqueda de acuerdo con los criterios establecidos en la ontología de visualización del interfaz del agente.

Marco obtiene una respuesta precisa a su consulta: un enlace a un artículo en el que se repasan las características, ventajas e inconvenientes de cada una de las especies de OWL; otro que le remite a un apartado concreto de una contribución a un congreso de *“Documentalistas en Internet”* que trata específicamente sobre las diferentes especies de OWL; y un último enlace, para ampliar información, le remite a la especificación de OWL publicada por el W3C.

Esta situación es hoy por hoy ciencia ficción, pero ya se están colocando los cimientos para que en un plazo no demasiado largo sea una realidad.

De hecho, es constatable que la actividad investigadora concerniente al proyecto Web Semántica es cada vez más intensa (el número y la importancia de los proyectos que se están llevando a cabo crecen constantemente) e implica a un mayor número de grupos y entidades de investigación (el “peso” investigador de estos grupos y entidades a nivel mundial es abrumador).

---

## 2.3 GRUPOS DE TRABAJO Y PROYECTOS SOBRE WEB SEMÁNTICA

Como vemos, la Web Semántica define una infraestructura cuyo objetivo es permitir el desarrollo de una plataforma universal para el intercambio de información. Esto supone definir un nuevo escenario del que la sociedad en su conjunto puede verse beneficiada, ya que los usuarios de la red podrán desarrollar sus tareas diarias e interactuar con la Web de una manera más amigable. En esta nueva Web los usuarios podrán, por ejemplo, disfrutar de nuevas posibilidades para la interrelación con otras personas mediante el establecimiento de redes sociales, les resultará más cómodo y seguro realizar transacciones comerciales con las empresas, o podrán gestionar más fácilmente cualquier trámite burocrático con las administraciones digitales, ya que dispondrán de mejores medios para proteger de una manera eficaz su privacidad (definiendo, por ejemplo, reglas de privacidad o preferencias que permiten un mayor nivel de control de los flujos de información).

No obstante, el éxito del modelo de Web Semántica según Euzenat (2002) se ve condicionado en gran medida por la satisfacción de cuatro requerimientos básicos que son críticos para su desarrollo:

- Es necesario el desarrollo de lenguajes de representación del conocimiento, mecanismos de identificación de recursos, e infraestructuras semánticas basadas en ontologías para la localización y acceso a los recursos identificados.
  - La heterogeneidad de los diferentes modelos semánticos que existen en la Web debe ser abordada desde el desarrollo de un modelo multicapa donde se definan diferentes lenguajes de representación modulares con los que establecer una infraestructura que favorezca la interoperabilidad entre modelos y la reutilización de recursos.
  - La precisión de los resultados dependerá de la definición de un método de razonamiento seguro y del desarrollo de nuevos modelos de
-

procesamiento que permitan el establecimiento de relaciones de confianza y el intercambio de pruebas de autenticidad.

- Por último, es imprescindible facilitar la implantación del modelo garantizando a los usuarios la disponibilidad de recursos y haciendo su funcionamiento transparente. De hecho, el éxito del modelo será completo en el momento en que la gente deje de referirse a la Web Semántica como tal y la denomine, sencillamente, la Web.

Afortunadamente, en la actualidad ya se están dando los primeros pasos para propiciar la implantación y el desarrollo de la Web Semántica, y multitud de instituciones y organismos están plenamente comprometidas con el desarrollo de proyectos y aplicaciones prácticas cuyo eje central es el uso de estas tecnologías. Desde organismos públicos como la Comisión Europea, a universidades o empresas privadas de primer nivel como Oracle, Nokia, IBM, Yahoo! o HP, muestran su interés, su compromiso y su confianza en los beneficios que el desarrollo del proyecto Web Semántica puede reportar al ámbito científico y académico, al industrial, al comercial, o al gubernamental (en definitiva, el beneficio que puede reportar a toda la sociedad en su conjunto), y ya disponen de laboratorios que se dedican en exclusiva a la investigación y al desarrollo en esta área.

Tras unos primeros años titubeantes en los que el proyecto era admirado por sus posibilidades al tiempo que observado con recelo por la inmadurez de las tecnologías sobre las que se basa, estamos entrando en un esperanzador periodo en el que determinadas barreras mentales están siendo derribadas, y donde los esfuerzos de inversión en investigación y desarrollo están dando sus primeros frutos tangibles.

En este apartado vamos a hacer un repaso por algunos de los grupos de trabajo y proyectos más importantes en esta área, distinguiendo entre aquellos que se enmarcan dentro de las actividades organizadas por el W3C y aquellos que están

---

patrocinados por instituciones ajenas al consorcio, y que a día de hoy permanecen activos.

### **2.3.1 GRUPOS Y PROYECTOS DEL W3C**

Dentro del W3C existe un área específica dedicada al desarrollo de la Web Semántica, denominada *Semantic Web Activity*, cuya misión consiste en aunar los esfuerzos de diferentes entidades y particulares en el desarrollo del proyecto Web Semántica y sus tecnologías (como los vocabularios RDF, SPARQL, GRDDL -(Halpin, Davis 2007)- y OWL), en hacer un seguimiento y publicitar las actividades que la organización lleva a cabo a nivel mundial en esta área.

Se divide en 8 grupos de trabajo que desempeñan diferentes actividades de coordinación, investigación y desarrollo en varios dominios. Veámoslos con algo más de detalle.

#### A) SEMANTIC WEB COORDINATION GROUP

Este grupo es el encargado de coordinar el resto de grupos que integran el área de Web Semántica del W3C (*Semantic Web Activity*) a través de un foro de debate donde se definen las interrelaciones e interdependencias que existen entre ellos, y donde se definen estrategias para evitar la duplicación o solapamiento de trabajos, y la fragmentación de proyectos por el desarrollo de tecnologías o vocabularios incompatibles entre sí.

#### B) RULES INTERCHANGE FORMAT WORKING GROUP

La misión de este grupo es desarrollar el núcleo de un lenguaje de reglas (y extensiones del mismo) que permita la traducción de reglas entre otros lenguajes de reglas heterogéneos y de esta forma poder ser utilizados y compartidos entre sistemas basados en reglas. The Working Group will have to balance the needs of a community diverse including Business Rules and Semantic users Web specifying extensions for which it can articulate a consensus design and which are sufficiently motivated by use cases. Actualmente el grupo está

---

trabajando en la definición del formato de intercambio de reglas o RIF (*Rule Interchange Format*) [22], un formato que en principio está diseñado para garantizar la interoperabilidad entre lenguajes de reglas en general, pero que por sus características particulares (como por ejemplo la utilización de referencias IRI's [23] o la definición de la interacción de las reglas RIF con RDF) para aquellos que se utilizan en la Web en particular.

#### C) OWL WORKING GROUP

La misión de este grupo de trabajo es refinar y extender la recomendación de OWL con un conjunto de elementos y propiedades que han sido identificados por usuarios y desarrolladores como necesarios para mejorar la capacidad expresiva y la facilidad de implementación del vocabulario. Dentro de este activo grupo existen diferentes proyectos cuyo objetivo es el desarrollo de aplicaciones que se apoyan para su funcionamiento sobre ontologías OWL. Dentro de las actividades de este grupo encontramos varios proyectos y aplicaciones:

- a) El proyecto BioPax [24] está desarrollando un formato común de intercambio de información en sistemas de información especializados en biología utilizando OWL.
  - b) El proyecto de automatización de informes técnicos del W3C [25] permite mantener de forma automática un índice en formato RDF de informes técnicos y especificaciones, y dispone de una serie de herramientas que permiten operar con esta información para facilitar la creación de este tipo de documentos.
  - c) El sitio web del proyecto MINDSWAP [26] está considerada como el primer sitio web semántico. Desarrollado en un principio utilizando el lenguaje ontológico SHOE [27], en la actualidad utiliza OWL para generar tanto las páginas del sitio, como las webs personales de los miembros del grupo de investigación.
-

- d) *SchemaWeb* [28] es un directorio que contiene esquemas RDF y ontologías OWL que pueden ser consultadas por los usuarios bien realizando búsquedas mediante palabras clave o explorando tripletas RDF utilizando un formulario que permite diseñar consultas más complejas. A su vez, dispone de una serie de servicios web que permiten a agentes software operar con el conocimiento almacenado.
  
- e) También dentro de este grupo se ha desarrollado el buscador semántico *Swoogle* [29]. Utilizando un interfaz muy similar al de, popular motor de búsqueda Google, permite a los usuarios consultar más de 10.000 ontologías, documentos y entidades descritos o que contienen metadatos definidos tanto en RDF como en OWL.

#### D) SEMANTIC WEB DEPLOYMENT WORKING GROUP

Este otro grupo tiene como tarea redactar informes técnicos sobre temas relacionados con el desarrollo y la implementación de RDF para generar nuevos vocabularios (como Skos Core -(Miles, Brickley 2008)-), el uso que se hace de OWL, o la integración de RDF en documentos XHTML (utilizando el vocabulario RDFa (Adida, Birbeck 2007), que aprovecha el poder expresivo de RDF para enriquecer la descripción de los recursos utilizando una serie de atributos nativos de XHTML (Pemberton et al. 2002) y otros propios.

#### E) SEMANTIC WEB INTEREST GROUP

Este grupo se ha concebido como un foro de debate donde tiene cabida la opinión tanto de miembros del W3C como de gente ajena a este organismo para discutir ideas, innovaciones y nuevas aplicaciones de la Web Semántica. También se tratan temas relacionados con tecnologías de apoyo que complementan y utilizan la Web Semántica, y la relación de estas con otras actividades del W3C así como con el contexto social y legal en el que se enmarca la propia Web.

---



Las principales actividades del grupo RDF Calendaring Taskforce se centran en el desarrollo de proyectos cuyo objetivo es el diseño de aplicaciones de agenda y calendario usando RDF (como la conversión de ficheros iCalendar de Apple a RDF [30], o la creación del esquema RDF para recursos de calendario [31]). Este grupo además mantiene foros abiertos para discutir sobre SPARQL (Prud'hommeux, Seaborne 2008) y FOAF (Brickley, Miller 2005).

#### F) SEMANTIC WEB HEALTH CARE AND LIFE SCIENCES INTEREST GROUP

Existe un grupo especializado en el área de las ciencias de la vida y la salud cuyo objetivo es la implantación de las tecnologías de Web Semántica para mejorar la colaboración, la investigación, y la innovación en este campo, principalmente en el desarrollo de sistemas de toma de decisiones. Entre los proyectos más destacados desarrollados dentro de este grupo encontramos los siguientes:

- a) BioDash [32]: Es un prototipo dirigido a la industria farmacéutica que permite asociar diferente información distribuida en diferentes servidores de bases de datos sobre enfermedades, compuestos farmacológicos, estadios de evolución de fármacos, biología molecular y marcadores moleculares. Para ello se propone describir la información utilizando tecnologías de Web Semántica para aprovechar todo este potencial de conocimiento sin necesidad de tener que diseñar nuevos modelos de datos. Utiliza la herramienta de interfaz Haystack (ver apartado 2.3.2) para generar “lentes” que permiten agregar y visualizar de diversas maneras información sobre marcadores moleculares y farmacológicos extraídos de diferentes fuentes.
  - b) SIMILE [33] es un proyecto conjunto entre las bibliotecas y el laboratorio de informática e inteligencia artificial (CSAIL) del MIT (*Massachusetts Institute of Technology*) que se centra principalmente en mejorar la interoperabilidad entre conjuntos de recursos digitales; esquemas, vocabularios u ontologías; metadatos y servicios disponibles en las
-

bibliotecas digitales utilizando para ello tecnologías semánticas. Aunque en un principio SIMILE está enfocado a su uso en el área de ciencias de la vida, los problemas que encuentran estos investigadores son muy similares al de otras comunidades, por lo que los resultados de este proyecto podrían ser fácilmente extrapolables a otras áreas científicas y servir de ejemplo sobre cómo las tecnologías de Web Semántica pueden ayudar a los investigadores proporcionándoles mecanismos simples y efectivos para integrar, gestionar y compartir información en la Web.

- c) AGFA on 'Connected Knowledge' [34]: Proyecto financiado por AGFA (proveedor de maquinaria sanitaria) para el desarrollo de un sistema de toma de decisiones que utiliza diferentes ontologías para mejorar la calidad de los servicios sanitarios. El sistema genera de forma dinámica rutas de decisión basadas en la información clínica que de forma continuada proporciona el paciente, en el entorno y en las pautas definidas en el sistema. A partir de estas se crea nueva información que permitirá evaluar y determinar los siguientes pasos a seguir.
- d) El proyecto *Active Semantic Documents* (ASD) [35] tiene como objetivo la creación de un sistema de toma de decisiones para profesionales de la medicina que permita reducir errores de diagnóstico, mejorando la eficiencia de los profesionales y la satisfacción de los usuarios con respecto a la praxis de estos profesionales. Las tecnologías de Web Semántica ayudan a conseguir estos objetivos al proporcionar las herramientas precisas para generar las ontologías sobre las que se basa el sistema, para la anotación de recursos y para el procesamiento de reglas. Los documentos ASD normalmente están descritos en formato XML y anotados utilizando una o más ontologías OWL que proporcionan el modelo conceptual para interpretarlos y razonar con ellos. La validación del contenido de los documentos se realiza de manera automática y dinámica aplicando reglas SWRL y consultas con RDQL (Seaborne 2004) sobre las anotaciones semánticas de los recursos.

- e) El proyecto *Knowledge Discovery Platform* de la compañía *Language and Computing* [36] pretende facilitar a los investigadores del área de la biomedicina interpretar y analizar tanto los vastos repositorios de literatura científica como otras fuentes de información complementaria. Para ello han desarrollado un sistema que de forma automatizada obtiene información de diferentes recursos científicos, taxonomías especializadas, bases de datos públicas y privadas, y de ontologías externas para integrarla en una única y enorme base de conocimiento. Sobre esta base de conocimiento diferentes motores de inferencia pueden descubrir relaciones complejas entre objetos y conceptos (e incluso representarlas gráficamente).
- f) El grupo *Partners Healthcare System's Clinical Knowledge Management*, una división de la compañía *Clinical Informatics Research and Development* [37], centra su actividad en la implementación de un sistema de gestión de conocimiento y toma de decisiones en sistemas de información clínica. Las ontologías y reglas sobre las que se basa el sistema ofrecen apoyo al profesional en tareas, por ejemplo, de observación o interpretación de datos clínicos. El proyecto incluye el desarrollo de editores de ontologías y reglas que se pueden integrar en sistemas de gestión de documentos.

#### G) POWDER WORKING GROUP

El grupo POWDER (*Protocol fOr Web DEscription Resources*) [38], de reciente creación, trabaja en el desarrollo de un protocolo para la publicación de descripciones de recursos web (como por ejemplo, metadatos sobre dichos recursos) utilizando el protocolo HTTP y vocabularios semánticos como RDF, OWL.

---

#### H) SEMANTIC WEB EDUCATION AND OUTREACH INTEREST GROUP

El grupo *Semantic Web Education and Outreach Interest* (SWEO) tiene como objetivo recopilar proyectos o prototipos basados en la implementación exitosa de las tecnologías de Web Semántica en organismos ajenos al W3C (ya sea en la empresa pública o en la privada) para proporcionar información sobre ellos y publicitarlos entre el público en general.

Existen diferentes sistemas implementados operativos, y otros no operativos en fase de prueba con prototipos (Herman 2007). A continuación repasamos los más importantes.

#### SISTEMAS IMPLEMENTADOS

- a) *The Semantic Web for the Agricultural Domain, a Semantic Navigation of Food, Nutrition and Agriculture Journal*. Food and Agriculture Organization of the United Nations (FAO) (Italia): Este proyecto se centra en el desarrollo de una ontología (AGRIS) y un tesoro multilingüe (AGROVOC) sobre agricultura, así como de las herramientas para la consulta de ambos.
  - b) *Real Time Suggestion of Related Ideas in the Financial Industry*. Bankinter (España): Es un sistema de gestión de conocimiento corporativo en el que los usuarios son invitados a introducir ideas o comentarios. Estas ideas son categorizadas de forma semiautomática mediante una equiparación semántica entre conceptos. De esta manera se favorece la recuperación de estas ideas por parte otros usuarios de la compañía.
  - c) *A Semantic Web Content Repository for Clinical Research*. Cleveland Clinic (Estados Unidos): Este proyecto pretende generar un repositorio sobre enfermedades cardiovasculares que se retroalimenta con nueva información extraída de los historiales de los pacientes, que utiliza una ontología OWL y un conjunto de reglas para realizar inferencias sobre
-

esta información, y dispone de un interfaz de búsqueda que gestiona consultas semánticas generadas en SPARQL.

- d) *Digital Music Archive*. ESIS y NRK (Noruega): El objetivo de este proyecto consiste en digitalizar toda la producción fonográfica de la radio televisión nacional noruega (NRK). Las tecnologías de Web Semántica se utilizan para mejorar el acceso a los recursos (muchos de ellos entradas de catálogo originalmente en formato papel).
  - e) *An Ontology of Cantabria's Cultural Heritage*. Fundación Marcelino Botín, (España): El proyecto consiste en el desarrollo de un portal web que incluye información y recursos digitalizados sobre el patrimonio cultural cántabro, que dispone de herramientas de navegación y un motor de búsqueda mejorado basados en una ontología.
  - f) *Prioritization of Biological Targets for Drug Discovery*. Eli Lilly (Estados Unidos): Este sistema, basándose en la integración de grandes cantidades de información de diferentes fuentes y utilizando diferentes ontologías (propias y externas), permite a los investigadores identificar indicadores biológicos para el descubrimiento de medicamentos.
  - g) *Geographic Referencing Framework*. Ordnance Survey (Reino Unido): Es un servicio de información geográfica que integra diversas fuentes de información que se apoya en varias ontologías (sobre hidrología, geografía política, o lugares y edificaciones), y que consigue de esta forma una integración de datos más eficiente y una clasificación de recursos más precisa.
  - h) *Composing Safer Drug Regimens for the Individual Patient using Semantic Web Technologies*. PharmaSURVEYOR Inc. (Estados Unidos): El sistema utiliza una ontología para ayudar a buscar el mejor régimen de medicación para un caso determinado de acuerdo a los síntomas o signos que presenta el paciente. Además el sistema puede presentar varias
-

soluciones posibles de acuerdo a la valoración de diferentes factores de riesgo.

- i) *Public Health Situation Awareness, School of Health Information Science.* Universidad de Texas (Estados Unidos): Producto de este proyecto de investigación, se ha desarrollado la aplicación SAPPHIRE (Situational Awareness and Preparedness for Public Health Incidences and Reasoning Engines). Es un sistema de vigilancia biomédica (epidemias, bioterrorismo, alertas veterinarias, etc.) que integra información de múltiples fuentes, y que permite monitorizar el desarrollo situaciones de crisis que pueden afectar a la salud pública. El análisis de la información obtenida mediante el uso de diferentes motores de inferencia puede permitir a las autoridades sanitarias competentes evaluar la situación y reaccionar forma rápida y eficiente. El sistema se desarrolla sobre una arquitectura multi-agente distribuida que utiliza OWL y RDF(S) para representar taxonomías médicas (como UMLS [39] de la National Library of Medicine o LOINC Logical Observation Identifiers Names and Codes [40]), y dispone tanto de mecanismos para clasificar enfermedades, como de herramientas para visualizar y navegar por la información de la que dispone.
  - j) *Improving the reliability of Internet search results using Search Thresher.* Segala (Irlanda): Dentro de este proyecto se ha desarrollado una aplicación que permite procesar los resultados de las consultas a un buscador web añadiendo contenidos adicionales en formato RDF con objeto de ofrecer al usuario información adicional que le ayude a valorar de una manera más simple la fiabilidad de los recursos obtenidos.
  - k) *Enhancing Content Search Using the Semantic Web.* Siderean Software y Oracle Corporation (Estados Unidos): La tecnología *Oracle Technology Network* (OTN) permite agregar recursos dispersos a través de un único portal web. Para ello utiliza una taxonomía que permite enriquecer la
-

descripción de estos recursos, y una ontología con la que es posible reagruparlos, presentarlos de una manera más intuitiva y realizar búsquedas más específicas. Incluye mejoras sobre las capacidades de búsqueda, navegación, interacción, y recuperación utilizando el vocabulario semántico de consultas SPARQL.

- l) *The sWoRDFish Metadata Initiative: Better, Faster, Smarter Web Content*. Sun Microsystems, Inc., y Zepheira (Estados Unidos): Esta aplicación permite la gestión del sitio web de *Sun Microsystems* mediante una serie de herramientas que permiten crear, clasificar y gestionar recursos y las relaciones que establecen entre ellos. Tiene la capacidad de importar metadatos de otras fuentes y de generar páginas dinámicas que incluyen listados de informaciones o servicios relacionados.
  - m) *Use of Semantic Web Technologies in Natural language interface to Business Applications*. Tata Consultancy Services Limited (India): Como parte del proyecto se ha desarrollado una aplicación de gestión empresarial con la que pueden interactuar los usuarios a través de un interfaz que gestiona consultas en lenguaje natural. Se apoya sobre una ontología que facilita la recuperación de datos y conceptos.
  - n) *Semantic Content Description to improve Discovery*. Vodafone, (Reino Unido): Integra en un portal web la descripción en formato RDF de varios productos (como tonos, juegos, etc.) de diferentes proveedores. A consecuencia de esta innovación y desde su puesta en funcionamiento el portal ha registrado un aumento significativo de las descargas de estos productos con descripciones enriquecidas.
  - o) *An Intelligent Search Engine for Online Services for Public Administrations*. Ayuntamiento de Zaragoza (España): Este grupo de trabajo español ha desarrollado un motor de búsqueda de servicios públicos ofertados por el Ayuntamiento de Zaragoza. Las búsquedas se basan en el procesamiento
-

de consultas en lenguaje natural y realizando inferencias utilizando ontologías.

- p) *Semantic-based Search and Query System for the Traditional Chinese Medicine Community*. Universidad de Zhejiang y Academia de Ciencias de Medicina China (China): Han desarrollado un sistema que integra un gran número de bases de datos relacionales sobre medicina tradicional china añadiendo un nivel de procesamiento semántico. Dispone de una herramienta con un interfaz gráfico que permite representar las bases de datos utilizando una ontología especializada. Las consultas al sistema se realizan a través de formularios.

#### PROTOTIPOS

- a) *Reduction of Errors in Radiological Procedure Orders*. Agfa Healthcare (Bélgica): Este prototipo dispone de un conjunto de pautas extraídas de diferentes fuentes OWL, a partir de las cuales, y utilizando diversos motores de razonamiento, es capaz de generar recomendaciones sobre los procedimientos radiológicos más adecuados que deben utilizar los técnicos clínicos para desarrollar su trabajo. De esta manera es posible prevenir (o al menos disminuir) errores de diagnóstico.
- b) *B2B Integration with Semantic Mediation*. BT Research y Venturing (Reino Unido): El objetivo de este prototipo es servir a las cadenas de suministros mediante la implantación de sistemas de apoyo operacional (OSS) que utilizan descripciones semánticas de los interfaces y mensajes que genera el sistema. De esta manera se garantiza la integración de diferentes proveedores de servicios web con los OSS. Gracias al uso de ontologías es posible reutilizar servicios y, de esta forma, reducir costes.
- c) *Active Knowledge Management for Integrated Operations*. Computas AS y Statoil (Noruega): Esta aplicación se ha desarrollado utilizando un motor de búsqueda semántico basado en ontologías que permite extraer y
-



procesar datos empíricos (así como otra información de interés) almacenados en una base de conocimiento para planificar prospecciones petrolíferas marinas.

- d) *Enhancing Web Searches within the Principality of Asturias*. CTIC Foundation, Universidad de Oviedo y Gobierno del Principado de Asturias (España): Dentro de este proyecto se ha desarrollado un portal para buscar información de carácter gubernamental (como leyes, notificaciones, etc.), utilizando SKOS Core y ontologías para facilitar al usuario la búsqueda de documentos jurídicos (aunque no esté familiarizado con la jerga y el vocabulario jurídico).
  - e) *Drug Ontology Project for Elsevier*. Elsevier, Aduna, y Vrije Universiteit (Holanda): El objetivo de este proyecto es crear un interfaz único para acceder a información de diferentes fuentes especializadas en el área de las ciencias de la vida. El sistema se basa en la descripción de recursos utilizando metadatos y en la utilización de un tesoro RDF como herramienta auxiliar.
  - f) *Ontology-Driven Information Integration and Delivery*. Chevron Information Technology Company (Estados Unidos): Este proyecto está formado a su vez por otros seis subproyectos que utilizan tecnologías de Web Semántica y ontologías para la gestión eficaz de información sobre ingeniería (química, petrolífera y gasera) y ofrecer servicios mejorados de búsqueda, navegación, recuperación y filtrado de recursos en la Web.
  - g) *Applying the Semantic Web to Internal Compound Repurposing*. Pfizer Ltd. (Reino Unido): Este sistema permite recuperar y reutilizar datos de trabajos y experimentos previos realizados en los campos de la medicina y la farmacia que son almacenados en formato XML y reutilizados exportándolos a formato RDF.
-

- h) *Semantic Web Technologies in Automotive Repair and Diagnostic Documentation*. Renault (Francia): La empresa Renault ha desarrollado un sistema basado sobre un modelo OWL para el diagnóstico de averías en automóviles que genera dinámicamente manuales de reparación en formato RDF, favoreciendo así el intercambio de información. El sistema además se vale del *feedback* proporcionado por el ingeniero para mejorar su rendimiento.
  
  - i) *Helping New Judges Answer Complex Legal Questions*. Consejo General del Poder Judicial (España): Iniciativa del consejo General del Poder Judicial que ha desarrollado un sistema de preguntas frecuentes (*frequently asked questions* o FAQ) mejorado con el uso de ontologías, que recoge jurisprudencia e información de interés proporcionada por más de 400 jueces de España.
  
  - j) *WEASEL, Vodafone R&D Corporate Semantic Web*. Vodafone Group Research & Development (España): WEASEL es un sistema de gestión de conocimiento corporativo que anota de forma automática recursos de acuerdo a una ontología sobre la que se apoya, e incluye un interfaz para realizar búsquedas en lenguaje natural.
-

### 2.3.2 GRUPOS Y PROYECTOS NO AUSPICIADOS POR EL W3C

Otros proyectos importantes que se están desarrollando fuera de los grupos de la *Semantic Web Activity* (y que no están listados por el grupo de interés *Semantic Web Education and Outreach*) son los siguientes:

a) Dentro del proyecto *Haystack* [41], y en activa colaboración con el proyecto SIMILE, se han desarrollado diferentes aplicaciones que utilizan tecnologías de Web Semántica y tienen como objetivo permitir a cualquier individuo personalizar de una forma ágil y sencilla la gestión de la información que utiliza de forma cotidiana. Entre las aplicaciones desarrolladas dentro de este proyecto destacamos:

- El explorador semántico *Haystack* [41] (que recibe su nombre del proyecto homónimo) es capaz de recuperar cualquier tipo de recurso de una manera flexible permitiendo, por ejemplo, definir diferentes criterios de búsqueda, ordenación o presentación de datos, y mejorando de esta manera la capacidad de cualquier individuo para encontrar la información que precisa.
  - *PLUM (Personal Lifetime User Modeling)* [42] es una plataforma software para capturar las necesidades a largo plazo de los usuarios (representándolas en formato RDF) para aplicarlas en la personalización de la interacción de estos con el sistema. *PLUM* está diseñado para servir como interfaz directamente en las aplicaciones y sistemas operativos más comunes de hoy día
  - *JOURKNOW* [43] es una aplicación que permite asociar a pequeñas piezas de información dispersas (como notas, conversaciones o comentarios) información contextual adicional como tiempo, espacio, interlocutores, etc. De esta forma se facilita su gestión y recuperación.
-

- Dentro del proyecto *Relo* [44] se han desarrollado una serie de *plugins* basados en RDF para la plataforma Eclipse [45] (una plataforma de software libre para el desarrollo de aplicaciones) que ayudan a los desarrolladores a explorar y comprender de una manera más sencilla bloques de código. Estos *plugins* permiten al programador crear diagramas intuitivos que le permiten centrarse en fragmentos de un bloque de código específico y descubrir relaciones relevantes entre elementos de los mismos. Así se consigue facilitar al programador una tarea tediosa como la edición de código, proporcionándole un entorno de trabajo más claro y ordenado.
  - *Piggy Bank* [46] es una extensión del navegador web *Firefox* que permite agregar información de diferentes sitios web y presentarla al usuario como un nuevo documento que puede ser almacenado localmente para ser compartido o para su posterior consulta por el propio usuario.
  - *Solvent* [47] es una herramienta complementaria de *Piggy Bank* que extrae información textual de un recurso web (ya sea texto plano o un documento HTML sin marcado semántico) y la transforma en un documento RDF que puede interpretar *Piggy Bank* (de alguna forma lo que hace es separar el contenido de las restricciones formales de los documentos y re-expresarlo de una manera más rica).
  - El motor de búsqueda Re:Search [48] permite enriquecer las consultas que realizan los usuarios agregando resultados relevantes obtenidos en consultas similares realizadas con anterioridad.
- b) *OntoWeb* [49] es una red formada por investigadores, científicos y profesionales de la industria y el comercio, provenientes de diferentes áreas, cuyo objetivo es desarrollar una plataforma basada en ontologías para el intercambio de información que permita facilitar la gestión del conocimiento y el comercio electrónico. Para ello fomentan la creación de grupos de
-

trabajo interdisciplinares y procuran el fortalecimiento de la influencia de organismos europeos en los procesos de estandarización de vocabularios semánticos basados en RDF y XML.

- c) Diferentes grupos de trabajo han desarrollado motores de búsqueda semánticos. Aparte de los ya mencionados Swoogle y Re:Search podemos destacar otros buscadores como *Semantic Web Search* [50] (un motor de búsqueda que localiza recursos basados en vocabularios RDF y soporta consultas SPARQL), y *CORESE*, un proyecto del *Institut National de Recherche en Informatique et en Automatique* francés que permite realizar búsquedas sobre grafos RDF [51].
  - d) *SWSI (Semantic Web Services Initiative)* [52] es un iniciativa dirigida a investigadores (tanto académicos como del campo de la industria) que estén involucrados en proyectos financiados por el DARPA y la Unión Europea sobre Web Semántica, o servicios web semánticos. Su objetivo es combinar estas tecnologías y fomentar su uso para impulsar el desarrollo de la Web Semántica a nivel mundial.
  - e) El proyecto *SemIPort (Semantic Methods and Tools for Information Portals)* desarrolla métodos y herramientas para crear y mantener portales semánticos de información para comunidades científicas. Se utilizan técnicas propias de áreas como los metadatos, la gestión y representación del conocimiento, la personalización, o la visualización de información. *SemIPort* define una infraestructura que permite realizar consultas a repositorios de documentos representados en formato RDF, pone a disposición del usuario herramientas para facilitar la interpretación y gestión de la información recuperada, y permite anotar documentos de manera tanto automática como manual.
  - f) *DBpedia* [53] es un proyecto que pretende hacer posible consultar la Wikipedia como si fuera una base de datos semántica. Wikipedia es la enciclopedia libre más grande de la Web. No obstante las posibilidades de
-

realizar búsquedas están bastante restringidas ya que las consultas se limitan a equiparar palabras clave con el texto completo de los artículos, imposibilitando, por ejemplo, recuperar artículos relacionados entre sí. Las tecnologías de Web Semántica posibilitan definir consultas más expresivas e interrelacionar diferentes recursos provenientes de diferentes fuentes. DBpedia extrae información estructurada de la Wikipedia y la hace semánticamente accesible expresándola en formato RDF. Actualmente, DBpedia dispone de información sobre más de 1.95 millones de “entidades” (personas, lugares, películas, etc.) que forman entre sí unos 103 millones de piezas de conocimiento (o tripletas RDF).

- g) DBin [54] es una aplicación semántica que proporciona a los usuarios expertos herramientas para crear “grupos de discusión” donde la comunidad de usuarios puede anotar y compartir cualquier recurso de su interés. Estas anotaciones están expresadas en RDF y el intercambio de información se hace sobre una arquitectura P2P, pero de manera transparente al usuario.
  - h) NeOn [55] es un proyecto financiado por la Comisión Europea en el que también participan otras instituciones líderes en Europa en el área del modelado del conocimiento y del desarrollo de ontologías. El objetivo del proyecto es crear una infraestructura (con su metodología asociada) para el desarrollo de servicios semánticos viables económicamente que tengan una aplicación real en el mercado. El proyecto pretende establecer una arquitectura idónea para la integración de ontologías, crear de un conjunto de herramientas para la gestión de las mismas, y definir una metodología para el desarrollo de aplicaciones semánticas.
  - i) KW (*Knowledge Web*) [56]: es un proyecto pretende fortalecer el sector de la industria y el comercio electrónico en Europa aplicando ontologías y otras tecnologías de Web Semántica. El proyecto incluye planes de formación, investigación y desarrollo para promocionar actividades y proyectos centrados en la implantación de tecnologías y servicios web semánticos en
-

los sectores industrial y comercial. De esta forma se intenta garantizar un impacto a largo plazo de las mismas que se pueda traducir en la implantación exitosa de sistemas semánticos en estos sectores. Además el consorcio de organismos que conforman KW realiza tareas de coordinación entre las diferentes comunidades interdisciplinarias involucradas en el proyecto para evitar duplicación o fragmentación en los trabajos que estas lleven a cabo.

- j) El proyecto ESSI (*European Semantic System Initiative*) [57] es la suma de diferentes proyectos cuyo principal objetivo es fomentar y fortalecer la investigación y la industria europeas a través de la estandarización a nivel mundial. Cada proyecto está especializado en un aspecto concreto de la Web Semántica como, por ejemplo, el establecimiento de la infraestructura básica del proyecto; el desarrollo y explotación de las tecnologías basadas en la Web Semántica; el enriquecimiento de las descripciones de los servicios web con metadatos semánticos; o el apoyo activo al proceso de transición de las tecnologías de ontologías desde el mundo académico a la industria. Dentro de los proyectos que se definen dentro de ESSI encontramos, por ejemplo, el proyecto Triple Space Communications (*TripCom*) [58] que tiene como objetivo añadir descripciones semánticamente accesibles a recursos, procesos y servicios, o el proyecto SUPER (*Semantics Utilised for Process management within and between Enterprises*) [59] cuyo fin es proporcionar métodos y herramientas para representar la información de manera que sea interpretable por máquina y de esta forma mejorar la gestión de los procesos de negocio a través de una única tecnología que integre ambas visiones. Otra iniciativa en la que participan conjuntamente el proyecto ESSI, el proyecto *Knowledge Web* [56] y la organización *DERI International* [60], es en la creación del instituto STII (*Semantic Technology Institute International*) [61]. Este instituto está concebido como una asociación interdisciplinaria colaborativa de diferentes grupos del ámbito científico, industrial o gubernamental involucrados en el proyecto Web Semántica y en el desarrollo de servicios web semántico, cuyo objetivo es convertirse en una de las
-

organizaciones de referencia dentro el campo de este campo. Uno de sus para convertir estas tecnologías en el eje central de la ingeniería de la computación moderna. Sus objetivos específicos consisten en definir nuevos modelos de negocio y mejorar la forma en que los clientes se comunican con las empresas.

- k) El proyecto NEPOMUK (*Networked Environment for Personal Ontology-based Management of Unified Knowledge*) [62] es otra iniciativa del grupo de investigación SMILE (*Semantic Information Systems and Language Ingeneering Group*) que tiene como objetivo crear una infraestructura semántica de código abierto sobre la que se puedan integrar componentes de diferentes desarrolladores para poner a disposición de los usuarios herramientas que les permitan gestionar su información, intercambiarla y compartirla dentro de una comunidad más amplia de usuarios (lo que comúnmente se conoce como *Semantic Desktop*). Específicamente, el proyecto pretende implantar este tipo de plataformas en el ámbito de las organizaciones y corporaciones, de manera que los trabajadores puedan mejorar la gestión de la información que utilizan para desarrollar sus actividades, y mantener una comunicación y colaboración activa con otros trabajadores con los que están relacionados (formando redes sociales).
  
  - l) SIOC (*Semantically-Interlinked Online Communities*) [63] es una iniciativa que pretende la integración de la información que circula en una comunidad web mediante el uso de una ontología web que recoge los principales conceptos y propiedades necesarios para identificar y describir en el ámbito de la Web Semántica la información generada por miembros de una comunidad web (como wikis, weblogs, etc.). La información está representada en formato RDF y ya existen varias aplicaciones que utilizan esta ontología (de hecho está en proceso de estandarización por el W3C).
  
  - m) JeromeDL [67] es un interesante proyecto patrocinado conjuntamente por la biblioteca de la Universidad tecnológica de Gdansk y DERI International [60]
-



que consiste en el desarrollo de una biblioteca digital semántica. La infraestructura semántica basada en ontologías sobre la que se sustenta el sistema facilita a sus usuarios el acceso a todos los recursos de la colección independientemente del formato de estos (incluido el fondo antiguo digitalizado y documentos en otros formatos no textuales, como presentaciones Flash). El sistema almacena e indiza los recursos formatos (a texto completo cuando es posible) y les asocia una descripción bibliográfica metadata enriquecida que se genera de manera asistida mediante el uso del módulo *JeromeAdmin*. Este módulo asimismo permite establecer protocolos de seguridad para contenido protegido (para evitar su copia o impresión), como las listas de control de acceso (ACL), y cargar los recursos en la base de datos de la biblioteca. La descripción básica obtenida puede a su vez ser complementada con descripciones adicionales del registro en formato MARC21 (con sintaxis XML), BibTex, o con otras anotaciones. Para acceder a los recursos se ponen a disposición del usuario herramientas que le permiten hacer *browsing* por materias y palabras clave, y realizar búsquedas tanto en los campos de la descripción enriquecida de los recursos como a texto completo; y para filtrar y contextualizar los resultados se aplican técnicas de desambiguación de términos (mediante operadores booleanos y el uso de vocabularios específicos). Además, permite hacer la colección de recursos accesible a otras bibliotecas digitales a través de un servicio web basado en el protocolo de bibliotecas extensible (ELP). De cada usuario se mantiene un perfil que es utilizado para mejorar la precisión de los resultados en consultas posteriores. El sistema los actualiza anotando información de carácter estadístico mientras el usuario navega por la base de datos. El formato de visualización de los resultados por defecto es XHTML aunque, siempre que sea posible, a los recursos se les asocia una plantilla XSL:FO, de manera que se da opción al usuario a elegir el formato de salida del recurso (PDF, RTF, XHTML, etc.). Otra aportación novedosa de JeromeDL con respecto a otros sistemas es su capacidad para aprovechar la perspectiva social de las

---

bibliotecas (digitales) para establecer redes de usuarios que pueden compartir opiniones sobre los recursos que consultan. Cada usuario tiene la opción de crear su propia lista de recursos favoritos (en los que, por ejemplo, puede anotar comentarios, opiniones, recomendaciones, etc.) para posteriormente compartirla con el resto de usuarios de la biblioteca. De esta manera cada recurso se convierten en un “comentario de blog” potencial al que otros usuarios pueden responder añadiendo sus propias anotaciones, generándose así debates y nuevo conocimiento.

- n) MARCONT [64] es una aplicación dirigida a su uso por parte de grupos de trabajo que permite el desarrollo cooperativo de ontologías. El portal pone a disposición de los usuarios un editor de ontologías con capacidad para generar y modificar clases, propiedades, espacios de nombre y registrar sus diferentes versiones, permite la traducción de unas ontologías en otras mediante reglas de conversión, y dispone de herramientas para la visualización gráfica de ontologías y la descripción de servicios web semánticos. Además, el proyecto se complementa con una ontología específicamente diseñada para su uso en bibliotecas digitales que combina descripciones MARC21 con metadatos Dublin Core y que puede integrar otras ontologías de manera que la experiencia de búsqueda del usuario mejore sustancialmente. Está íntimamente ligada al proyecto Jerome-DL, el cual pretende convertirse en una plataforma de evaluación de la ontología MarcOnt.
  
  - o) El proyecto Calais [65] pretende posibilitar el uso de aplicaciones semánticas mediante un servicio web de generación de metadatos a partir del contenido de piezas enviadas por los usuarios. Estos metadatos se incrustan en el contenido que mandan los usuarios y posteriormente es sometido a un proceso de categorización del contenido en entidades (personas, lugares, organizaciones, etc.), hechos (relaciones entre entidades) y eventos (relaciones entre entidades contextualizadas temporal o topográficamente).
-

Los metadatos se almacenan de manera centralizada y se les asigna un identificador GUID (*Globally Unique Identifier*) [66]. Calais es un sistema extensible que admite la introducción por parte de los usuarios de sus propios metadatos de forma que complementan y matizan los generados de manera automática. Estos metadatos pueden ser utilizados para construir mapas conceptuales, redes sociales, o enlazar recursos relacionados, permitiendo de esta forma, por ejemplo, mejorar la navegación por un sitio web, proporcionar información contextualizada, u organizar, compartir y filtrar contenidos.

Como vemos, indistintamente organismos públicos y empresas tanto de Estados Unidos como de la Unión Europea (con el patente liderazgo del W3C y la especialmente entusiasta participación de países como Alemania, Gran Bretaña o España), se han postulado como los principales valedores de la Web Semántica y están fomentando la puesta en marcha de diferentes proyectos cuyo fin último es desarrollar una plataforma sólida para el intercambio de información (ya sea en el ámbito académico, científico, industrial, financiero o comercial) sobre la que sea posible implementar aplicaciones y servicios web basados en tecnologías semánticas.

El utópico escenario que dibujaba Berners-Lee en los trabajos en los que definía por vez primera la Web Semántica, parece que ya no lo es tanto (o al menos se está más cerca de los planteamientos de partida). Otra prueba de ello son las cada vez más numerosas herramientas software que facilitan la creación de recursos y aplicaciones de Web Semántica.

---

## **2.4 APLICACIONES SOFTWARE PARA EL DESARROLLO DE LA WEB SEMÁNTICA**

A pesar de que la Web Semántica se encuentra en las primeras fases de desarrollo ya existen multitud de recursos descritos en vocabularios semánticamente accesibles (RDF, FOAF, RSS, etc.) y cada vez más se desarrollan pequeñas ontologías verticales (especializadas en un dominio muy específico) para resolver problemas concretos utilizando estándares como OWL. No obstante, hasta hace poco tiempo no existían apenas herramientas para editar, visualizar, o localizar recursos RDF u ontologías de una manera eficiente, o para desarrollar aplicaciones semánticas completas.

En la actualidad su número se ha multiplicado y se están comenzando a desarrollar los primeros sitios web y portales que funcionan sobre una infraestructura ontológica que les permite aprovechar las posibilidades de búsqueda y acceso a recursos relevantes que ofrecen las técnicas de recuperación semántica de información.

En este apartado vamos a hacer un repaso (que pretende ser más ilustrativo que exhaustivo) de algunas de las herramientas software a las que, a día de hoy, pueden recurrir los desarrolladores para crear recursos y aplicaciones de Web Semántica. Para mejorar la comprensibilidad de la revisión, hemos clasificado las herramientas en 9 categorías principales.

### **A) EDITORES RDF**

Como hemos visto en apartados anteriores, uno de los principales problemas que obstaculiza una rápida implantación del modelo de Web Semántica es la dificultad que existe para anotar semánticamente los recursos de la Web de una manera sencilla e intuitiva (un problema común para todos los creadores de contenido, ya sean usuarios comunes o profesionales). Por esta razón, son

---

muchas las energías dirigidas a desarrollar editores que minimicen el esfuerzo del usuario, bien sea mediante ayudas contextuales o a través de interfaces gráficas amigables. De hecho, la mayoría de editores de instancias RDF y ontologías que existen hoy día combinan ambas soluciones para facilitar en la medida de lo posible la tarea de crear un modelo RDF y evitar al usuario al tecleado de código en bruto.

Entre los editores RDF más conocidos encontramos Protégé [68], RDFAuthor [69], o IsaViz [70].

Algunos de estos editores RDF disponen de módulos que permiten específicamente el desarrollo de ontologías (principalmente mediante interfaces gráficas). Además del ya citado Protégé, otros editores de ontologías destacados son Ontolingua [71], SWOOP [72], SweDE [73], Ontoedit (Horrocks, Hendler 2002), o SMORE [74]. También existen soluciones comerciales como Ontostudio [75] (orientado para su uso en la empresa) o Altova Semantic Works [76].

Otro tipo de editores de metadatos son los anotadores, herramientas permiten generar metadatos de una manera automática o semiautomática a partir del contenido de recursos web de los que se extraen una serie de conceptos de interés de acuerdo a una ontología de apoyo. Ejemplos de este tipo de aplicaciones son el anotador semántico SWAN (*Semantic Web Annotator*) [77] o Magpie [78].

## B) HERRAMIENTAS DE VISUALIZACIÓN

Debido a la naturaleza del modelo de datos RDF que permite representar instancias en forma de grafos, las herramientas gráficas y de visualización pueden ser de gran ayuda para el diseño y la edición de recursos RDF/OWL, por lo que muchos editores incluyen este tipo de ayudas para facilitar dichas tareas. Ese es el caso de editores como Isaviz, o Protégé que utiliza una serie de *plug-in's* como Jambalaya, OWL plugin u Ontoviz para visualizar recursos RDF y

---

ontologías OWL. Otros programas que también utilizan representaciones gráficas de este tipo de estos recursos son RDF-Gravity [79], que permite visualizar varios documentos RDF simultáneamente y realizar búsquedas a texto completo sobre ellos; RDF Room [80], con el que es posible visualizar grafos RDF desde una perspectiva isométrica, o el propio servicio de validación de documentos RDF del W3C [81] que genera gráficos sencillos a partir de un trozo de código RDF.

### C) SOFTWARE DE ALMACENAMIENTO Y RECUPERACIÓN

En esta categoría hemos incluido aquellas aplicaciones que sirven específicamente para almacenar recursos RDF y que facilitan el acceso a los mismos mediante un lenguaje de consulta semántico (estándar o propietario). Ejemplos representativos de este tipo de aplicaciones son la plataforma *Sesame* [82], que dispone de herramientas tanto para crear repositorios RDF (a los que es posible acceder mediante consultas definidas con *SerQL*), como para realizar inferencias sobre RDF(S); 3Store [83], un repositorio desarrollado en C y MySQL; *Kowari* [84], que permite almacenar y analizar recursos RDF y OWL, y que utiliza como lenguaje de interrogación el vocabulario Tucana [85]; el sistema de almacenamiento distribuido y recuperación YARS (*Yet another RDF Store*) [86], el repositorio ARC2 [87] desarrollado sobre PHP, o el repositorio RDFStore [88] que soporta consultas con SPARQL.

### D) PLATAFORMAS DE DESARROLLO

Existen también diferentes plataformas de programación que integran herramientas de para crear aplicaciones software basadas en tecnologías semánticas. La más relevante es Jena [13] una infraestructura construida sobre Java que proporciona un entorno de programación para RDF(S) y OWL. Soporta la formulación de consultas SPARQL y dispone de un motor de inferencia basado en reglas. Otra plataforma representativa, desarrollada en C, es Redland [89].

---

Dispone de varias librerías para el procesamiento, almacenamiento, consulta y recuperación recursos RDF. La plataforma cuenta además con interfaces de programación de aplicaciones (API) para otros lenguajes de programación como Perl, PHP, Python y Ruby.

A parte de estas plataformas, la gran mayoría de los principales lenguajes de programación disponen de sus propias librerías desarrolladas específicamente para trabajar con RDF y OWL. Así encontramos, por poner solo algunos ejemplos, RAP [90] y pOWL [91] para PHP; Perllib [92] para Perl; o Pyrple [93] y RDFLib [94] para Python.

#### E) NAVEGADORES

También es posible encontrar navegadores web específicos (o complementos que se instalan en los navegadores tradicionales) que permiten visualizar recursos RDF. Entre ellos encontramos aplicaciones como Tabulator [95], que se instala como complemento de los navegadores Firefox u Ópera, y que permite editar y visualizar datos y documentos RDF de manera integrada en el propio navegador. Otros ejemplos relevantes son los navegadores desarrollados dentro de los proyectos Haystack [96] (que incluye funcionalidades para la gestión de fotografías y archivos multimedia) y SIMILE Longwell (que permite realizar búsquedas facetadas)[97]. También existen navegadores más especializados que están diseñados para visualizar exclusivamente recursos FOAF, como FOAFnaut [98] y FOAF Explorer [99].

#### F) PAQUETES DE GESTIÓN DE ONTOLOGÍAS

En este grupo vamos a incluir aplicaciones que están diseñadas específicamente para trabajar con ontologías.

---

Las hay que, como Ontobuilder (Engels, Lech 2003) y Text-to-Onto (Maedche, Voltz 2001), permiten extraer los conceptos relevantes de un corpus de recursos web para posteriormente generar una ontología, abarcando de esta manera las primeras fases del ciclo de vida de las ontologías.

No obstante, si lo que necesitamos es una plataforma que defina un entorno en el que poder llevar a cabo la gestión integral de una ontología (incluyendo desde el proceso de diseño y edición, hasta su agregación, mantenimiento o refinamiento) debemos recurrir a herramientas más potentes como PROMPT [100], Chimaera (McGuinness 2000) o IBM Web Ontology Manager [101].

#### G) PLATAFORMAS DE DESARROLLO DE SERVICIOS WEB SEMÁNTICOS

Este tipo de aplicaciones proporcionan una infraestructura y una serie de herramientas que permiten describir y publicar este tipo de servicios sobre el soporte de ontologías de procesos. Un ejemplo de este tipo de software es el paquete Internet Reasoning Service (IRS) [102]. Dispone de tres módulos funcionales principales: un servidor en el que se almacenan las descripciones de los servicios web, incluyendo funciones, resolución de problemas y modelos de dominio; un módulo de publicación que enlaza los servicios a sus descripciones correspondientes, y un cliente que soporta la invocación de los servicios web.

#### H) MOTORES DE INFERENCIA

Otra categoría de herramientas semánticas son los razonadores o motores de inferencia. Son aplicaciones que toman como entrada los asertos definidos en una ontología e infieren nuevos asertos a partir de estos. Pueden ser utilizados, entre otras cosas, para:

- Descubrir relaciones jerárquicas entre clases.
  - Determinar la existencia de conceptos (o elementos) sinónimos.
-



- Detectar definiciones de clase inconsistentes.

Dado que para definir ontologías semánticamente operativas es necesario definir las utilizando OWL DL u OWL Lite (tienen una menor capacidad expresiva que las ontologías definidas en OWL Full, pero garantizan un razonamiento eficiente), la gran mayoría de los motores de inferencia que se desarrollan tienen capacidad para razonar utilizando lógica descriptiva y reglas definidas en SWRL. Existen múltiples soluciones de código abierto como Pellet [103], FaCT ++ (*Fast Classification of Terminologies*) [104], KAON2 [105], el motor de inferencia integrado en el paquete Jena [13], o CWM (*Close world Machine*) [106], un razonador que permite realizar consultas, transformar y filtrar información contenida en recursos serializados como RDF/XML o RDF/N3. También es posible encontrar aplicaciones comerciales como es el caso de RacerPro [107], que además de gestionar ontologías OWL puede ser utilizado como motor de inferencia por editores semánticos como, por ejemplo, Protégé.

## I) APLICACIONES WEB

En esta categoría hemos incluido aplicaciones que están a disposición de los usuarios de la Web y que suponen los primeros intentos serios de acercar la Web Semántica de una manera tangible al público en general.

- PORTALES SEMÁNTICOS

Son sitios web cuyo objetivo es ofrecer información especializada de áreas tan diversas como el comercio, la salud, la cultura, el e-learning o el gobierno electrónico, a una comunidad cerrada de usuarios que demandan herramientas de búsqueda y navegación precisas y versátiles para obtener recursos de su interés. Las tecnologías de Web Semántica, y en concreto el uso de ontologías verticales, permiten desarrollar este tipo de herramientas, maximizando la reutilización de recursos y facilitando la contextualización y personalización de resultados.

---

En la red es posible encontrar multitud de portales semánticos, aunque sin duda el más relevante es el portal SWED [108], enmarcado dentro del proyecto SWAD-Europe, y que puede ser considerado como el primer portal semántico de código abierto de la Web. Mantiene un directorio de instituciones y organizaciones relacionadas con la naturaleza y biodiversidad construido sobre una infraestructura ontológica.

Para el desarrollo de portales semánticos existen varias alternativas. Una de ellas es recurrir a plataformas específicas que facilitan el diseño de este tipo de sitios web, como es el caso de Ontoviews [109] (aplicación utilizada, por ejemplo, en la creación del portal semántico del gobierno de Finlandia [110]). Otra posibilidad consiste en adaptar portales semánticos ya existentes a un dominio determinado para satisfacer necesidades concretas de información. Este es el caso del portal SWED, cuyo código y herramientas han sido utilizados como modelo para desarrollar multitud de otros portales temáticos.

- BUSCADORES SEMÁNTICOS

Al igual que en la Web basada en HTML, la Web Semántica también necesita de servicios como los motores de búsqueda que exploren e indiquen la Web para facilitar la localización y dar acceso a sus recursos a través de un interfaz que sea sencillo de utilizar. Existen motores verticales, que realizan búsquedas sobre dominios muy específicos utilizando ontologías especializadas (como es el caso del buscador Kango [111] que permite planificar viajes realizando un análisis semántico de las demandas de los usuarios) y motores de búsqueda horizontales o generalistas.

Prácticamente cualquier sitio web construido utilizando tecnologías semánticas dispone de su propio motor de búsqueda vertical para facilitar el acceso a sus recursos, por lo que su variedad y número es más que considerable. Sin embargo, el caso de los motores de búsqueda

---

horizontales es algo más complejo ya que su fin último es indicar y proporcionar acceso a todos los recursos la Web. Por esta razón el desarrollo de este tipo de buscadores está aún dando sus primeros pasos, aunque ya existen herramientas más que interesantes.

Una de las primeras iniciativas para desarrollar un buscador semántico fue Swoogle [29] un motor que dispone de más de 10.000 ontologías y 2,3 millones de documentos RDF indizados, representados tanto con sintaxis XML, como con N-Triples o N3. De hecho funciona más como un repositorio de documentos de Web Semántica que como un motor de búsqueda tradicional. Permite realizar búsquedas sobre ontologías, instancias, términos y documentos que incluyen fragmentos en formato RDF/XML, así como realizar *browsing* sobre los resultados para descubrir recursos relacionados. Otra característica de este buscador es que archiva las diferentes versiones que puedan existir de un mismo recurso para facilitar a los usuarios realizar búsquedas retrospectivas.

Otros ejemplos de este tipo de aplicaciones son *Onthology* [112], un repositorio de ontologías centralizado (definido como una antología de ontologías) que utiliza el esquema OMV para describir, clasificar, dar acceso, evaluar y almacenar ontologías (a disposición de los usuarios de manera libre o bajo diferentes licencias de uso); y *Oyster* [113], otro repositorio (pero en este caso descentralizado) que permite la gestión, búsqueda, e intercambio de ontologías utilizando redes P2P.

En la línea de estos buscadores encontramos *Sindice* [114], un índice de documentos de Web Semántica (recursos RDF, ontologías, etc.) generado a partir de técnicas de agrupamiento o *clustering* que permite localizar recursos a través de palabras clave, URIs o IFPs (*Inverse Functional Property*) (búsqueda sobre propiedades con un valor unívoco, como el correo electrónico), aunque no ofrece soporte para realizar consultas sobre

---

tripleas RDF. Aplicaciones similares son SWSE [115], Watson [116], Falcons [117], y Ontogator [118].

Otros buscadores ofrecen experiencias de búsqueda más parecidas a la que proporcionan los buscadores que actualmente encontramos en la red, como es el caso de Hakia [119], cuyo objetivo es ofrecer resultados de calidad frente a resultados con un alto nivel de popularidad que ofrecen buscadores como Google. La calidad de los resultados se valora de acuerdo a tres criterios: i) credibilidad de la fuente (por lo general fuentes verticales recomendadas por bibliotecarios libres de sesgo comercial), ii) actualidad de la información y iii) relevancia del recurso. Por lo tanto, es un motor que puede ser de especial utilidad para profesionales de campos específicos frente al usuario que realiza búsquedas poco especializadas. Hasta el momento el motor ofrece cobertura de fuentes autorizadas para dominios como medicina, finanzas, justicia, ciencia, viajes, arte e historia, y tiene indizados recursos principalmente en inglés, aunque también ofrece documentos en español, portugués, italiano y polaco (progresivamente se irán añadiendo más idiomas). Además, introduce funcionalidades sociales adoptadas de la Web 2.0 como poner a disposición de los usuarios un foro en el que individuos interesados en un mismo tema pueden intercambiar información e impresiones.

También podemos encontrar soluciones comerciales que desarrollan buscadores avanzados desarrollados sobre arquitecturas multi-agente. Es el caso de *Semantic Web Search* [120], una plataforma para el desarrollo de aplicaciones de Web Semántica y agentes que dispone de un motor de búsqueda que permite a cualquier usuario (humano o agente software) realizar consultas sobre tripleas RDF utilizando vocabularios como RSS, Dublin Core o FOAF. Como resultado, el sistema devuelve el listado de los recursos en los que se ha producido una equiparación exacta.

---

Por su parte, el motor de búsqueda Zitgist [121] ofrece como principal novedad un interfaz que simplifica el proceso de generar consultas semánticas. Para ello los usuarios deben describir las características de aquello que está buscando. En primer lugar el usuario debe elegir el sujeto que el motor debe buscar como, por ejemplo, personas, grupos, documentos, foros de discusión o lugares geográficos. A continuación debe describir las características del sujeto seleccionado seleccionando una serie de propiedades y definiendo sus valores (el número de estas características varía según se esté utilizando el interfaz básico o el avanzado) (Giasson 2007). Este tipo de búsqueda soporta la definición de relaciones semánticas complejas del tipo *foaf:knows*, y el uso de ayudas gráficas interactivas (*widgets*), como mapas para especificar lugares geográficos, o calendarios para definir fechas.

Un enfoque diferente es el que propone Powerset [122], un buscador que permite a los usuarios realizar consultas en lenguaje natural. Para conseguir este objetivo el motor realiza un análisis lingüístico y contextual profundo de las frases contenidas en los recursos de la Web a partir del cual genera un índice semántico de hechos sobre cada frase.

Otra propuesta alternativa, y quizá más cercana a los hábitos de búsqueda del usuario actual, es la que propone Yahoo!. Microsearch [123] es un buscador híbrido que supone el puente de unión entre la Web de hoy día y la Web Semántica ya que combina los resultados de la búsqueda tradicional con metadatos RDF, RDFa y microformatos extraídos de los recursos web (en un futuro aceptará el uso de GRDDL).

- APLICACIONES WEB 2.0

La Web Semántica no es ajena al fenómeno de la Web 2.0 y el software social ya que de hecho para algunos autores este movimiento supone la transición natural del actual modelo de Web al modelo semántico.

Una escenificación clara de esta relación es que una de las primeras

---

aplicaciones semánticas operativas y abiertas al público en general es un sitio web de redes sociales. Twine [124] se define como un servicio para etiquetar recursos que permite organizar, compartir y descubrir información relacionada con los intereses del usuario dentro de una red de usuarios con intereses similares que se agrupan de acuerdo a diferentes criterios (amistad, comunidades especializadas en un tema, grupos de trabajo, etc.). Basado en tecnologías de Web Semántica, Twine organiza automáticamente la información aportada por cada usuario, la analiza e infiere cuáles son sus intereses para poder generar recomendaciones que van ganando en precisión con el tiempo gracias a un proceso de aprendizaje.

- APLICACIONES DE *DESKTOP* SEMÁNTICO

Otra herramienta interesante es el sistema de código abierto Gnowsis [125], una aplicación en versión beta que añade interfaces semánticas a diferentes aplicaciones de escritorio comunes como buscadores, programas de ofimática y estadística, o visores de documentos, unificándolas. Permite a los usuarios interrelacionar entre sí diferentes recursos (documentos, fotografías, correos electrónicos, contactos, etc.), anotarlos, realizar búsquedas y navegar entre ellos. De esta forma los recursos almacenados en un ordenador personal pueden formar una pequeña Web Semántica a escala.

El sistema dispone de un servidor local (con capacidad para agregar información heterogénea y distribuida, e indizarla a texto completo), un buscador semántico, y unas aplicaciones o *plug-ins* denominadas “*adaptadores*”, que transforman los recursos obtenidos de las diferentes aplicaciones de escritorio (exploradores de sistema, editores de etiquetas mp3-ID3, gestores de correo, o buscadores web) en recursos RDF. De esta forma, la representación y acceso a la información se hace de una manera uniforme independientemente de la aplicación que la genera. La filosofía

subyacente a Gnowsis es la de mantener la estructura de los datos sin modificarla, y ampliar las aplicaciones de escritorio sin sustituirlas.

## **2.5 VIABILIDAD DEL PROYECTO WEB SEMÁNTICA**

A pesar de que el modelo teórico de la Web Semántica está aceptado hace tiempo por una gran parte de la comunidad científica, no han cesado de surgir voces críticas que, esgrimiendo diferentes argumentos, han puesto en entredicho su viabilidad.

Algunos autores se aproximan a la Web Semántica como una aplicación de la Inteligencia Artificial (IA) al ámbito web, y advierten sobre las inconsistencias semánticas que se pueden derivar de la definición de algunos elementos del modelo (Patel-Schneider, Fensel 2002), o de la dificultad de extraer semántica (más allá de la explícita) de los documentos expresados en lenguaje natural (Sparck-Jones, 2004). Sin embargo, y a pesar de estas deficiencias (que no tienen por que ser insalvables) es importante no perder la perspectiva de lo que un proyecto de esta envergadura puede llegar a ofrecer. Tomemos como ejemplo el caso de la propia IA. Es una rama de las Ciencias de la Computación cuyo principal (y ambicioso) objetivo es conseguir diseñar entidades artificiales que exhiban algún tipo de comportamiento inteligente, pero por el momento aún no se ha conseguido una computadora que sea capaz de emular fielmente el razonamiento humano. No obstante, diferentes áreas de la IA sí que tienen una aplicación real y sirven para resolver eficazmente determinados problemas. De igual manera, la Web Semántica pretende aportar su propia visión del mundo y proponer algunas soluciones.

Las perspectivas de éxito a corto plazo de la Web Semántica se ven mermadas debido a la gran dificultad que en la actualidad supone el etiquetado semántico de los recursos, tanto por parte de los usuarios como de los propios desarrolladores. Como apuntábamos anteriormente, no es posible pensar en una

---

Web poblada por agentes procesando y compartiendo información si antes no existen suficientes recursos anotados con los que operar.

La tarea de etiquetar semánticamente recursos de una manera escrupulosa puede llegar a convertirse en un trabajo intelectual titánico, que en apariencia no ofrece especiales ventajas ni recompensas acordes al esfuerzo empleado a aquel que lo realiza (ya que se hace de manera absolutamente altruista). Por esta razón, gran parte de los esfuerzos de investigación se están dirigiendo al desarrollo de mecanismos y aplicaciones que permitan el etiquetado de recursos (Soo et al., 2003) y la generación automática de ontologías (Tho, Cao 2006).

No obstante, hoy día en la implantación y realización del proyecto de Web Semántica tienen un mayor peso específico las propuestas que se hacen desde los movimientos de software libre que promueven la Web 2.0, que las que puedan provenir del campo de la IA. Este nuevo concepto de Web (del que aún no se tiene una definición formal consensuada) se caracteriza principalmente por adoptar un enfoque orientado al usuario, a la interacción y al desarrollo de redes sociales, donde los sitios web actúan más como puntos de encuentro entre usuarios que como Webs tradicionales (independientemente del uso de una tecnología en concreto).

Algunos ejemplos de aplicaciones Web 2.0 son Flickr [126], del.icio.us [127] y CiteUlike [128], donde los usuarios proponen sus propias categorías para describir fotos, listados de hiperenlaces favoritos o referencias bibliográficas, respectivamente, y las comparten de forma altruista con el resto de usuarios para facilitarles el acceso a esos recursos. Esta filosofía supone *de facto* la aplicación práctica, aunque a pequeña escala, de conceptos, principios y tecnologías que encajan dentro del proyecto Web Semántica. Así, por ejemplo, encontramos las *machine tags* [129] que se utilizan en el etiquetado social y cooperativo, vocabularios RDF, como FOAF (*Friend Of A Friend*) (Brickley, Miller 2005), RSS (*RDF Site Summary*) (Beged-Dov et al., 2001B)), o RDFa (Adida, Birbeck 2007), un vocabulario que permite la integración de metadatos RDF en



documentos XHTML utilizando un número reducido de atributos (algunos nativos de XHTML y otros nuevos).

De esta forma el creador de contenidos puede describir los documentos de una manera precisa y los agentes pueden interpretar esta información de una manera robusta. Admite la inclusión de elementos Dublin Core [130], FOAF (Brickley, Miller 2005), Creative Commons [131] y definiciones XML Schema (Fallside, Walmsley 2004). Otra iniciativa es eRDF (*embedded RDF*)[132], un vocabulario que define una sintaxis sencilla para escribir código HTML o XHTML, sin añadir nuevos elementos o atributos, de manera que dicha información puede ser extraída como RDF mediante un *parser* eRDF o mediante hojas e estilo XSLT, y que permite trabajar con el resto de tecnologías de apoyo de HTML como las plantillas CSS.

También es interesante el vocabulario GRDDL (Gleaning Resource Descriptions from Dialects of Languages) (Connolly 2007) que como su propio nombre indica es un mecanismo para recuperar descripciones de recursos de dialectos de lenguajes. GRDDL define un conjunto de etiquetas (que permiten marcar determinados bloques de información de un recurso XML para que sea extraída y procesada por aplicaciones semánticas), y la notación especial que es necesario incluir para que se active el mecanismo de extracción. De esta manera sería posible, por ejemplo, extraer descripciones en formato RDF desde el contenido de un fichero XHTML.

Sin embargo la opción más popular a día de hoy para incluir información semánticamente accesible en documentos web es la utilización de microformatos (Khare, Tantek 2006), una serie de vocabularios sencillos que utilizan atributos presentes en XHTML para añadir información semántica a páginas web en formato (X)HTML, RSS, Atom o XML mediante una serie de clases definidas, principalmente, con atributos *class* (aunque también se utilizan otros como *id*, *title*, *rel* o *rev*). Existen multitud de microformatos diferentes (muchos de ellos estándares abiertos) que ofrecen soluciones para resolver problemas

---

específicos de etiquetado semántico de recursos. Su uso supone un paso decisivo hacia la implantación de la Web Semántica ya que a su sencillez de uso, hay que añadir la posibilidad de procesar los recursos descritos con estos vocabularios utilizando otros vocabularios específicos de la Web Semántica como SPARQL o el ya comentado GRDDL.

Otras iniciativas son más ambiciosas y van más allá del simple marcado de recursos ya que proponen generar ontologías simples a partir de las semánticas emergentes que se pueden extraer del conjunto de etiquetas usado por los usuarios de una determinada comunidad (Naaman 2007) (Gruber 2005) (Mika 2005).

Por lo tanto, no sería descabellado afirmar que muchas de las posibilidades de éxito a corto plazo de la Web Semántica pasan por el éxito de las aplicaciones Web 2.0 (Hendler 2007). Esta necesidad de definir semánticas más simples está justificada por la gran dificultad que existe para organizar y representar el conocimiento de una forma homogénea y común para todos los dominios.

Sin embargo, la idea de desarrollar una Web semántica adquiere pleno sentido en determinados ámbitos dependientes de dominio donde el uso de ontologías compartidas y el análisis semántico de recursos puede suponer una ventaja evidente (Shadbolt, Hall, Berners-Lee 2006). Esta idea vuelve a hacer resurgir con fuerza el antiguo debate sobre la idoneidad de desarrollar dos redes paralelas no excluyentes entre sí: una Web comercial abierta al público en general y una Web académica orientada a dar servicio a los docentes e investigadores.

Esta última se llegó a plasmar en el irrealizado proyecto de diferentes instituciones gubernamentales y académicas para desarrollar Internet2 [133]. La idea no sería crear una Internet de primera clase y otra de segunda, sino ofrecer diferentes servicios acordes a las características y necesidades de los usuarios, y cada uno desde la plataforma adecuada (¿acaso acudimos a una biblioteca pública cuando estamos buscando un libro de ecuaciones diferenciales?).

---

Abundando en esta idea, es precisamente en el ámbito de las bibliotecas universitarias donde la Web Semántica podría encontrar un terreno abonado para su implantación, y una vía para penetrar de manera progresiva en otros ámbitos de la sociedad. Se dan varios factores que favorecen esta posibilidad:

- Por un lado, las bibliotecas universitarias están inmersas en un proceso de transformación hacia lo que se ha venido a denominar la “*biblioteca digital universitaria*”, donde las colecciones de recursos en papel pasan a ser repositorios de recursos en formato digital a los que la comunidad universitaria (docentes, investigadores y alumnos) puede acceder desde cualquier ordenador personal a través de las redes abiertas de información (García, García 2001).
- Las características específicas de su comunidad de usuarios y las necesidades de información de estos hace necesario el desarrollo de sistemas avanzados de recuperación y filtrado de información que no se pueden ofrecer con las técnicas y herramientas que existen en la actualidad.
- Aunque suelen ser centros multidisciplinares, las áreas de conocimiento que cubren las bibliotecas universitarias están bien delimitadas y podrían ser descritas mediante ontologías (siempre que existiese un grado de consenso lo suficientemente alto entre los expertos de dichas áreas).

En este contexto, las bibliotecas universitarias digitales desarrolladas sobre una infraestructura semántica podrían representar el papel de portales de entrada a las grandes redes de información científica (Euzenat 2001) ya que permitirían establecer:

- Una plataforma universal de intercambio de información científica de todo tipo (datos, procedimientos, intereses y filiaciones de personas, etc.), extensible, robusta frente a fallos e inconsistencias, y que garantice
-

la interoperabilidad y la autenticidad de sus recursos, así como el acceso seguro a estos.

- Mecanismos de gestión de información distribuida a gran escala que, a pesar del gran volumen de recursos de la red, permitan al usuario localizar y procesar información de forma transparente (incluso combinada y obtenida de múltiples fuentes distribuidas).
- Mecanismos de gestión de los flujos de trabajo para evitar la réplica de procesos de manera innecesaria.
- Mecanismos para compartir recursos distribuidos de forma coordinada, de manera que con el análisis exhaustivo de datos y el modelado predictivo sea posible localizar aquellos recursos que no son de utilidad en la red y descartarlos de forma transparente al usuario.
- Una infraestructura para poder desarrollar una ciencia cooperativa, donde los investigadores podrán, de una manera sencilla, diseminar sus trabajos, crear agrupaciones de recursos de interés (a modo de bibliografías recomendadas), contactar con otros investigadores para crear grupos de trabajo de su área de trabajo o multidisciplinares, o resolver problemas de forma conjunta.
- Servicios de control para controlar asuntos como, por citar solo algunos, la autoría, autenticidad, autenticación y evaluación de los recursos, o temas relacionados con la seguridad y la confidencialidad.

De esta manera, se pueden convertir en auténticos sistemas de gestión de conocimiento que integren de forma transparente la creación y el acceso al conocimiento en el entorno de trabajo de los investigadores y docentes. Podrían poner a disposición de sus usuarios los medios necesarios para detectar semánticas emergentes (a través de métodos de generación automática o semi-automática de ontologías), donde el acceso y presentación de la información se hará de forma contextualizada (por ejemplo, a partir de un organigrama

---

determinado o a una jerarquía de procesos) y personalizada de acuerdo al nivel de experiencia o a la tarea que debe desempeñar un usuario determinado.

Estas perspectivas y proyectos como JeromeDL [67] hacen que el maridaje entre bibliotecas universitarias y Web Semántica deje de verse como una mera hipótesis, ya que dentro del nuevo marco que dibuja el Espacio Común Europeo de Educación e Investigación se hace una apuesta decidida por apoyar la educación, la investigación y la docencia en las tecnologías de la información, y por convertir a las bibliotecas universitarias en elementos claves del proyecto.

Hemos comprobado que la Web Semántica avanza paliando las evidentes dificultades que un proyecto de tal envergadura conlleva y que es posible aplicar estos avances a multitud de posibles escenarios. En nuestro caso, vamos a acotar el contexto de investigación a las bibliotecas digitales.

---

## 2.6 LA WEB SEMÁNTICA EN EL CONTEXTO DE LAS BIBLIOTECAS DIGITALES

Las nuevas Tecnologías de la Información y la Comunicación han contribuido a la modernización de las bibliotecas, desde la automatización de los catálogos hasta poner a disposición de los usuarios las colecciones accesibles en línea, pero también han provocado la aparición de nuevos problemas (relacionados con nuevos formatos, con la gestión de sistemas distribuidos, con el desarrollo de normas de edición electrónica, de estándares bibliotecarios, de modelos de evaluación de calidad, con nuevas normas legales sobre propiedad intelectual, o con el diseño de interfaces de acceso a la información) ante los que hay que proponer nuevas soluciones.

Las primeras bibliotecas digitales se limitaban a evolucionar a partir de una biblioteca física tradicional y se centraban básicamente en proporcionar acceso en línea a los recursos que formaban su colección (la mayor parte de ellos revistas o libros en soporte papel convenientemente digitalizados), llegando de esta manera a un público más amplio (York, Studer 2005).

Este uso tan limitado de las nuevas tecnologías de la información ha supuesto en gran medida un desaprovechamiento de las múltiples posibilidades que estas ofrecen para optimizar el procesamiento, disseminación, recuperación y análisis de la información.

Estas primeras bibliotecas digitales han arrastrado el lastre de las bibliotecas tradicionales, tanto en su concepción como en la forma de gestionar las colecciones, aunque las diferencias entre ambos tipos de bibliotecas son más que evidentes. Tal es esta dependencia que hasta la misma denominación "*biblioteca digital*" está viciada en origen por varias razones.

Mientras que los recursos que forman las colecciones de las bibliotecas tradicionales se encuentran localizados físicamente en un edificio (de hecho, biblioteca etimológicamente significa lugar para guardar libros, del griego

---

“*biblion*”= libro, y “*thêke*”= armario), y es posible acceder a ellos de una manera tangible ya que, por lo general, se encuentran en un soporte físico que el usuario puede llevar consigo (papel, microfilm, etc.), en las colecciones digitales se trasciende el concepto de libro con la inclusión de recursos gráficos, sonoros y multimedia, y se trasciende el concepto de lugar físico ya que, por lo general, los recursos se encuentran deslocalizados en la Red.

Por lo tanto, y de una manera simplista, las bibliotecas digitales se podrían definir como sistemas de información que disponen de colecciones de hipervínculos seleccionados que dan acceso a los recursos en formato digital que conforman el fondo de la biblioteca.

Algunos autores, como el doctor en filología griega Alfonso Martínez Díez, denominan a estas colecciones de hiperenlaces *dictiografías* (del griego “*diktýon*”= red y “*graphía*”= escritura o representación gráfica) para diferenciarlas claramente de las bibliografías, que relacionarían al resto de recursos (independientemente de su formato) que no tienen presencia en Internet.

Esto nos lleva a plantearnos si no sería más correcto (o preciso) referirse a las bibliotecas digitales como *dictiotecas*, entendidas como “*lugares*” (que se corresponderían con los sitios web que sirven de plataforma al sistema de información) donde se almacenan colecciones de hiperenlaces.

Así, y tomando como base la definición de una de las acepciones de “*biblioteca*” recogidas en el diccionario de la Real Academia de la Lengua [2], y las definiciones que sobre “*biblioteca digital*” proporcionan Marchionini (2000), y Smeaton y Callan (2003), podríamos definir “*dictioteca*” como “*un sistema de información que alberga colecciones de hiperenlaces sobre recursos en diversos formatos, que son análogos o semejantes entre sí, ya por las materias de que tratan, ya por la época y nación o autores a que pertenecen, y que tiene asociados una serie de servicios para facilitar el acceso a estos recursos a diferentes comunidades de usuarios utilizando para ello diversas tecnologías*”.

---

No obstante, la redefinición de las bibliotecas digitales debería llevarse a un nivel más profundo que el puramente terminológico. Uno de los puntos clave en este proceso de reinención es la mejora de la eficiencia de procesos como el acceso y recuperación de información, ya que de ellos depende en gran medida la calidad del servicio que una biblioteca puede ofrecer a sus usuarios.

De hecho, la recuperación de información en la Web ha revolucionado las metodologías y técnicas utilizadas en la investigación y supone una nueva forma de acceder a los recursos de información.

Para la recuperación de recursos textuales se han desarrollado múltiples técnicas y algoritmos basados en diferentes modelos (vectorial, booleano, probabilístico) que permiten realizar búsquedas en el contenido de los mismos mediante la equiparación léxica de términos. Sin embargo, estas técnicas no resultan eficientes cuando tratamos con material no textual (debiendo recurrir a otras metodologías como el reconocimiento de formas o el análisis cromático, en el caso de los documentos gráficos), o cuando necesitamos acceder a información implícita en el contenido (lo cual requiere un proceso complejo de inferencia). Este último tipo de búsquedas solo sería posible definir las dentro de la Web Semántica.

Según Finin y colaboradores (2005) se definen tres enfoques diferentes sobre cómo puede llegar a ser la recuperación de información en la Web Semántica. Una primera aproximación es la que considera que la Web Semántica no variará sustancialmente de la Web que conocemos hoy excepto porque los documentos dispondrán de una descripción enriquecida *incrustada* en su código con metadatos comprensibles por máquina que permitirá a los agentes software determinar la semántica de su contenido.

Otra aproximación consistiría en definir descripciones semánticas externas sobre los recursos web convencionales a los que referencian. En ambas, las herramientas de búsqueda tradicionales podrían ser capaces de procesar asertos

---



RDF a modo de contenedores de palabras clave aplicándoles una serie mínimas modificaciones (Finin et al. 2005).

La última posibilidad que barajan consistiría en la creación de motores de búsqueda semánticos que indicen y recuperen exclusivamente recursos de Web Semántica frente a los buscadores tradicionales.

No obstante, e independientemente del enfoque, la búsqueda de información en la Web Semántica difiere de la búsqueda tradicional por varias razones (Finin, Ding 2006):

- La Web Semántica está orientada a ser interpretada por máquinas y las anotaciones semánticas de documentos están orientadas a ser el puente entre los usuarios y los motores de búsqueda semánticos.
  - La forma en que está codificado el conocimiento (con vocabularios que utilizan el modelo de datos RDF) difiere también tanto de la información poco estructurada presente en la mayoría de recursos web, como de la información altamente estructurada contenida en las bases de datos. Es información semi-estructurada que permite la definición de clases, propiedades, o restricciones sobre asertos (algo impensable en vocabularios más sencillos como XML), lo cual implica el desarrollo de técnicas de indización y recuperación más complejas que las tradicionales.
  - Los documentos semánticos pueden ser realmente complejos (incluyendo definiciones de clases, propiedades, hechos, reglas, y metadatos) por lo que el proceso de creación de estos supone un trabajo intelectual previo para determinar en qué forma y medida los motores de búsqueda van a poder razonar con estos documentos.
  - La estructura en forma de grafos de los documentos RDF difiere en gran medida de la de los documentos HTML, por lo que también serán diferentes las estrategias para el descubrimiento automático de
-

documentos de Web Semántica y las métricas para determinar su relevancia.

Las tecnologías de Web Semántica mantienen una relación simbiótica con las bibliotecas digitales. Por un lado aplicadas a las bibliotecas digitales básicamente permiten definir una descripción enriquecida de objetos y repositorios, y el establecimiento de esquemas comunes de conocimiento en forma de ontologías, que favorecen la interoperabilidad de los recursos, es decir, la capacidad de acceder de forma consistente y coherente a clases similares de objetos digitales y servicios distribuidos en repositorios heterogéneos (entre los que, en un futuro, se podrían incluir repositorios de información de carácter personal no privada, o de información corporativa).

Pero existen otras muchas aplicaciones de estas tecnologías en el ámbito de las bibliotecas digitales ya que pueden ofrecer soluciones, por ejemplo, para el desarrollo de interfaces de usuario, la mejora de la navegación, visualización y presentación de contenidos, el modelado de usuarios (obteniendo descripciones más ricas y precisas de las necesidades de información de los individuos), la personalización de servicios (tanto a nivel individual como a nivel colectivo en el caso de comunidades extensas de usuarios), o la mejora de la interacción de los usuarios con el sistema (York, Studer 2005).

---

### **2.6.1 LOS CENTROS DE RECURSOS PARA EL APRENDIZAJE E INVESTIGACIÓN: UN MARCO PARA EL DESARROLLO DE LA WEB SEMÁNTICA**

Dentro del ámbito universitario las bibliotecas siempre han jugado un papel predominante como fuente de recursos de información (principalmente libros y revistas científicas especializadas) para docentes, investigadores y estudiantes. No obstante, las necesidades específicas de esta comunidad de usuarios (información muy específica y actual) y las características de la propia literatura científica (en constante crecimiento exponencial y generalmente con un alto nivel de obsolescencia), hacían que la tarea de mantenerse informado de todas las novedades científicas de un área determinada fuera lenta, costosa e ineficiente.

No será hasta la segunda mitad del siglo XX cuando aparezcan las primeras bases de datos científicas, unas potentes herramientas que bien mediante acceso remoto o a través de la consulta local de soportes de almacenamiento masivo facilitan al investigador el acceso a recursos de su interés previo pago por cada documento o mediante suscripción. También en este periodo se produce la progresiva sustitución de los libros por las revistas científicas como fuente primordial de información, se empieza a imponer el formato electrónico al papel y aparecen las primeras bibliotecas digitales.

En este nuevo siglo, las bibliotecas universitarias (entendidas tanto en su dimensión física como lógica) están llamadas a cumplir un importante papel como centros intermediarios entre los usuarios y la información, permitiendo el acceso, difundiendo y generando nuevo conocimiento. Y lejos de ser una mera declaración de intenciones, este rol de las bibliotecas está reconocido institucional y normativamente en el ámbito de la Unión Europea dentro del escenario que dibuja el Espacio Común Europeo de Educación e Investigación (que incluye al Espacio Europeo de Educación Superior (EEES) y el Espacio Europeo de Investigación (EEI)), que materializa la convergencia europea en materia de educación superior. El desarrollo de este marco de actuación

---

pretende armonizar los sistemas de educación superior y generalizar un modelo de aprendizaje permanente y común en todos los países miembros de la UE.

Su implantación, que supondrá la adopción de nuevos métodos tanto de enseñanza y aprendizaje, como tecnológicos y de gestión, afectará de lleno a la concepción que hasta ahora se tenía de lo que es una biblioteca universitaria, pasando de ser meros centros proveedores de recursos para investigadores, docentes y estudiantes, a convertirse en Centro de Recursos para el Aprendizaje e Investigación (CRAI) [134], un modelo que toma como referencia los "*Learning Resources Centres* o *Curriculum Materials Centres*" británicos (Buttler, Tipton 1992). Estos centros son unidades de información orientadas al usuario y al objeto para el cual necesita información, y su misión consiste en asesorar, orientar y apoyar la investigación, la docencia y el aprendizaje.

En este contexto, las bibliotecas universitarias deben convertirse en auténticos agentes dinamizadores que, a modo de motor, impulsen la implantación del nuevo modelo educativo, integrando la perspectiva digital mediante la incorporación de las nuevas tecnologías de la información y la comunicación (TIC) en los servicios que ofrecen.

El centro debe adoptar un papel proactivo, intentando anticiparse a las necesidades de sus usuarios y ser capaz de ofrecerles información adaptada y personalizada, utilizando para ello mecanismos capaces de seleccionar, organizar y facilitar el acceso a los recursos que requieren, tomando en consideración tanto sus características y necesidades particulares a «largo plazo», como sus habilidades y/o nivel de conocimientos, y la forma en que estos evolucionan en el tiempo.

En esta doble dimensión de bibliotecas físicas y digitales, los CRAI deben cumplir determinados objetivos (Moscoso 2003), [134]:

- crear y diseñar una biblioteca digital a partir de las necesidades reales de los usuarios.
-

- evaluar y seleccionar los recursos que forman parte de su repositorio.
- describir, analizar y organizar los materiales digitales.
- proporcionar a los usuarios acceso a estos recursos.
- preservar la seguridad y la conservación de estos recursos e incrementar la visibilidad de los resultados de investigación.
- potenciar el aprendizaje y la colaboración en red con otros centros.
- desarrollar otras tareas de gestión.

Entre las herramientas básicas que los CRAI deben ofrecer a sus usuarios habría que incluir sistemas eficientes de recuperación y filtrado de información, que permitan proporcionar a cada individuo, de la forma más precisa y exhaustiva posible, aquellos recursos que son de su interés y que le van a permitir desarrollar las tareas o cumplir los objetivos que han provocado su petición de información. Por lo tanto, el desarrollo de sistemas de DSI (es decir, los sistemas de filtrado y recomendación), será una aportación fundamental para que los CRAI resulten una herramienta de utilidad para las diferentes tipologías de usuarios a los que da servicio, de forma que puedan extraer, estructurar y compartir su conocimiento (Gonçalves et al. 2004).

El modelo de CRAI está concebido como una plataforma universal para el intercambio de información dentro de un marco de cooperación interuniversitaria. Esto implica aunar esfuerzos por parte de estos centros para adoptar normas y estándares que permitan la compatibilidad de sistemas, la representación del conocimiento y la configuración de redes, de forma que sean centros «lógicos» independientes de plataforma.

El proyecto Web Semántica, por definición, asume estas premisas y dispone de las especificaciones y herramientas tecnológicas idóneas para desarrollar una infraestructura sólida sobre la que construir los CRAI. Como vemos, la relación que se puede establecer entre Web Semántica y los CRAI es absolutamente

---

simbiótica, puesto que los CRAI podrán aprovechar las capacidades que ofrecen las tecnologías de Web Semántica para representar y organizar el conocimiento y el modelo de Web Semántica encontrará en los CRAI un marco de implantación idóneo.

La adopción de la Web Semántica como nuevo paradigma tecnológico, no sólo implicará una profunda transformación en la Web que conocemos hoy día, sino que, en ámbitos como las bibliotecas o en disciplinas como la Documentación (cuyas materias primas de trabajo son la información y el conocimiento), su implantación podría suponer una revolución equiparable a la ocurrida al sustituir los catálogos manuales por los automatizados. Las diferentes tecnologías que subyacen en este nuevo modelo de Web ofrecerán al profesional de la información la posibilidad de disponer de un arsenal de herramientas flexibles y potentes que, de forma análoga a los lenguajes documentales, le permitirán ejercer el control necesario sobre la información, para procurar su descripción, acceso y recuperación eficaz, pero a unos niveles impensables hasta ahora.

En este nuevo escenario, los documentalistas tendrán que desempeñar un papel fundamental en el control de la descripción de recursos y en el diseño conceptual de ontologías, que servirán para el desarrollo de diferentes servicios y procesos dentro de los sistemas de información. Multitud de estos procesos, que se suelen realizar a diario de forma manual o semiautomática, se verán transformados de una forma radical. Y no solo nos referimos a la recuperación de información (tal y como hemos visto hasta ahora) ya que, por ejemplo, en el proceso de catalogación o descripción de recursos se tenderá a utilizar lenguajes para el modelado de datos (basados en RDF) que servirán como plataforma de descripción.

Gracias al uso de editores semánticos la gran mayoría de los recursos estarán ya «catalogados» en origen, antes de entrar al sistema de información. Progresivamente se impondrá el uso de perfiles de aplicación (espacios de nombre estándar que podremos combinar como deseemos de acuerdo a las

---

necesidades de nuestro sistema), al tiempo que otros formatos de descripción como MARC (*Machine-Readable Cataloguing*) se verán avocados a desaparecer y a ser sustituidos por otros estándares de descripción de recursos que emplean metadatos, como los estándares DCMI (*Dublin Core Metadata Initiative*) [130], PRISM (*Publishing Requirements for Industry Standard Metadata*) [135] o MODS (*Metadata Object Description Schema*) [136].

La automatización de catálogos ya ocasionó en su momento que los sistemas de clasificación clásicos, como la nada intuitiva CDU o los encabezamientos de materia, quedaran desplazados en ese nuevo contexto de trabajo. De forma similar, la indización en un entorno semántico se verá modificada, pasando a ser un proceso asistido por aplicaciones capaces de extraer términos representativos de cada recurso y sugerir otros extraídos de estructuras jerárquicas como ontologías web, taxonomías o tesauros.

Las tecnologías de Web Semántica abren también una nueva vía para encontrar solución a problemas endémicos en la generación de resúmenes automáticos, ya que en una Web donde los documentos estén descritos semánticamente, se podrán desarrollar agentes capaces de distinguir en diferentes tipos de recursos aquellas partes o áreas susceptibles de ser resumidas y, además, asistir en el análisis morfológico, sintáctico, semántico y pragmático del contenido, recurriendo para ello a las ontologías adecuadas.

Las capacidades que ofrecen estas tecnologías para compartir y reutilizar recursos permitirá a los sistemas de bibliotecas definir, por ejemplo, sus políticas de préstamo o incluso las de expurgo, especificando diferentes conjuntos de reglas que permitan controlar estos procesos a través de una red de agentes y ontologías compartidas.

Los nuevos sistemas de información aprovecharán su infraestructura semántica para aplicar técnicas de minería Web Semántica y buscar modelos o patrones en el contenido y estructura intrínseca de sus recursos, en las relaciones que establecen con otros recursos y en la forma en que son utilizados por los

---

usuarios. Esta información será, a su vez, de gran utilidad para el modelado de perfiles de usuario, donde aspectos mudables como las preferencias, costumbres de navegación, gustos, intereses, permisos de acceso a los recursos, etc., de un individuo concreto podrán ser expresados mediante metadatos y actualizados dinámicamente para reflejar fielmente sus variaciones en el tiempo. De esta forma se podrán desarrollar sistemas de filtrado y recomendación más precisos.

De nuevo puede parecer que estemos planteando meras hipótesis, pero no cabe duda de que si estas potentes tecnologías representan el futuro de los sistemas de información, las Ciencias de la Documentación (y sus profesionales) no pueden obviar esta circunstancia y deben apostar por adaptarse a esta nueva Web en la que, por sus propias características, el papel de los profesionales de la información puede ser más relevante que nunca. De hecho, el concurso de dichas tecnologías puede ser crucial en el desarrollo de productos y servicios de valor añadido para las bibliotecas digitales.

---





# SERVICIOS DE VALOR AÑADIDO PARA BIBLIOTECAS DIGITALES

### 3.1 INTRODUCCIÓN

Tal y como indican Marchionini (2000) y Tramullas (2002), las bibliotecas digitales son proveedores de información claramente orientados al usuario, y uno de sus principales objetivos es facilitar a sus usuarios el acceso a los recursos de las diferentes colecciones que conforman su fondo mediante servicios interactivos de valor añadido para el usuario final.

La variedad de servicios de valor añadido que puede ofrecer una biblioteca digital es enorme, pudiendo encontrar desde los tradicionales servicios de difusión selectiva de información, a servicios de aparición más reciente como los de redifusión de contenidos.

Estos servicios se basan en la personalización del acceso de acuerdo al perfil de los usuarios de la biblioteca, consiguiendo de esta manera satisfacer de una manera más precisa sus necesidades de información. Esto implica que en su diseño sea imprescindible un estudio previo de las necesidades específicas de la comunidad de usuarios a la que sirve.

Pero no solo el usuario se ve beneficiado ya que, según Barrueco (2007), para una biblioteca digital la implantación de este tipo de servicios le permite situarse en una posición de privilegio para competir con a otras bibliotecas digitales ya establecidas. Por esta razón, Tramullas (2007) afirma que el valor añadido es uno de los activos más importantes de la actividad bibliotecaria en las bibliotecas digitales.

---

Los procedimientos e indicadores que se pueden mejorar implementando servicios de valor añadido en un sistema de información (Taylor 1986) se pueden clasificar en seis categorías de acuerdo al problema que palian. Existen medidas específicas para reducir el ruido documental (es decir, disminuir en la medida de lo posible aquellos documentos recuperados por el sistema que no son relevantes para el usuario, y que por lo general se obtienen como resultado de una consulta demasiado genérica); medidas para mejorar la calidad de la información (valorando la precisión, la originalidad y la diversidad de los datos disponibles en el sistema); medidas para facilitar a los usuarios el uso del sistema y proveerles de herramientas de ayuda y orientación; medidas para mejorar la adaptabilidad del sistema a los cambios del entorno y de los requerimientos de los propios usuarios; y medidas para el ahorro de recursos. A continuación desglosamos cada clase pormenorizadamente:

- *Reducción del ruido documental*
    - Acceso a la información: La mejora del acceso se consigue mediante el desarrollo de tecnologías capaces de permitir al usuario disponer de información discriminada que le permite satisfacer sus necesidades de información específicas. Necesita de la identificación y descripción precisa de la información, así como del desarrollo de herramientas de búsqueda y localización eficaces.
    - Enlaces externos: Consiste en proporcionar al usuario vínculos con fuentes y sistemas de información externos para ampliar la oferta de información con colecciones y recursos no disponibles en nuestro sistema.
    - Precisión: Es la capacidad del sistema para extraer la fracción de documentos, de entre los documentos recuperados, que son relevantes para el usuario de acuerdo a las características particulares de su necesidad de información.
-

- Selectividad: Una adecuada selectividad permite al sistema recuperar documentos pertinentes a raíz de una interpretación correcta de la consulta formulada por un usuario.
  - *Calidad de la información*
    - Exactitud: Consiste en garantizar la transferencia sin errores de datos e información al usuario.
    - Exhaustividad: La capacidad del sistema para extraer de forma exitosa la fracción de documentos que son relevantes para el usuario de entre toda la colección de recursos del sistema. Por lo tanto, un alto grado de exhaustividad permite ofrecer al usuario la posibilidad de tener acceso a todos los recursos disponibles sobre un tema específico. No obstante, una alta exhaustividad influye negativamente en la precisión del sistema.
    - Actualidad: el valor agregado por la inmediatez de los datos suministrados por el sistema y por la capacidad del propio sistema para reflejar los enfoques actuales del pensamiento en su estructura, organización, y vocabularios para el acceso.
    - Confiabilidad: Este indicador se refiere al grado de confianza que inspira el sistema a los usuarios gracias a la fiabilidad y consistencia de su funcionamiento a lo largo del tiempo.
  - *Facilidad de uso, ayuda y orientación al usuario*
    - Facilidad de la exploración: Consiste en poner a disposición del usuario herramientas de navegación que le permitan encontrar información que le pueda ser de utilidad de forma casual mediante la técnica del ojeo.
    - Formato: Una estructuración, maquetación y presentación adecuada de la información supone un valor añadido que permite la navegación eficiente y sencilla por la información para encontrar aquellas que son relevantes para el usuario.
-

- Interfaces: El desarrollo de interfaces es una potente herramienta que permite al usuario comunicarse y obtener respuestas del sistema (además de servirle como medio para la comprensión del entorno de trabajo y para la adquisición de experiencia en su utilización), y a través del mismo, el sistema es capaz de interpretar las necesidades del usuario.
  - Accesibilidad física: Consiste en facilitar el acceso físico a la información del sistema.
  - Organización: La organización del conocimiento de acuerdo a un esquema determinado redundará en una mejor y fácil localización lógica de la información.
  - *Adaptabilidad*
    - Flexibilidad: Consiste en proporcionar al usuario diferentes maneras y enfoques que le permitan trabajar con la información disponible en el sistema en la forma más conveniente.
    - Simplicidad: Permite añadir valor al servicio de información mediante la presentación a los usuarios de la información recuperada de una manera clara y precisa.
    - Promoción: Consiste en desarrollar diferentes actividades con la finalidad de consolidar la presencia del sistema entre la comunidad de usuarios (incluidos los potenciales).
    - Correspondencia con la necesidad: Consiste en desarrollar las tareas necesarias para satisfacer las necesidades específicas de cada usuario mediante el conocimiento (lo más exhaustivo posible) de los mismos. Para ello es de utilidad el diseño de perfiles de usuario en los que se recoja, por ejemplo, sus preferencias, intereses personales y profesionales, nivel de conocimientos o habilidades.
-

- *Reducción de costos*
  - Ahorro de recursos: El valor añadido a la información se consigue mediante el diseño de un sistema de gestión y de toma de decisiones operativas dirigido al ahorro de recursos técnicos y económicos.

La mejora de uno o varios de estos indicadores o procesos permiten que las bibliotecas digitales alcancen objetivos de excelencia en el servicio que ofrecen a sus usuarios.

En este capítulo nos vamos a centrar en concreto en el análisis de dos servicios de valor añadido que posibilitan mejorar el funcionamiento de las bibliotecas digitales: los sistemas de filtrado y recomendación de información, y los servicios de redifusión de contenidos. También vamos a examinar algunas de las herramientas con las que es posible definir una infraestructura adecuada sobre la que desarrollar estos servicios: los sistemas de organización del conocimiento, las tecnologías de Web Semántica y el modelado lingüístico difuso.

## **3.2 LOS SISTEMAS DE FILTRADO Y RECOMENDACIÓN DE INFORMACIÓN**

Uno de los principales retos que hoy día tienen que afrontar los sistemas de información es la gestión eficaz del gran volumen de documentos que almacenan para poder facilitar a los consumidores de información el acceso a recursos que satisfagan sus necesidades de una manera sencilla y ágil. Esta necesidad se vuelve más acuciante en una sociedad como la actual donde el nivel de exigencia de los usuarios es cada vez mayor.

Tradicionalmente, en el ámbito de las bibliotecas el problema de la sobrecarga de información se ha abordado adoptando diferentes medidas, como por ejemplo la creación de servicios de difusión selectiva de información (DSI), en los que de acuerdo al perfil de los usuarios suscritos al servicio se generan periódicamente (o a petición del propio usuario) una serie de alertas en las que

se les notifica de la existencia de recursos que se adecuan a sus intereses (Aksoy et al. 1998) (Foltz, Dumais 1992).

Los servicios de DSI han sido objeto de estudio desde diferentes áreas de investigación, como en el ámbito de desarrollo de sistemas multi-agente (Decker, Sycara, Williamson 1997) (Kuokka, Harada 1995), y por descontado, en el de las bibliotecas digitales (Faensen et al. 2001). En la actualidad muchos servicios de DSI se implementan sobre plataformas web con una arquitectura multi-agente donde, por un lado, existe una serie de agentes intermediarios que realizan la equiparación de los perfiles de usuario con los documentos y, por otro, diferentes agentes de entrada y salida que se encargan de realizar las suscripciones al servicio y presentar las alertas a los usuarios (Altinel, Franklin 2000) (Carzaniga, Rosenblum, Wolf 2000) (Fabret et al. 2001) (Yan, García-Molina 1999). La información generalmente viene estructurada según un determinado modelo de datos y los perfiles de usuario se definen utilizando una serie de palabras clave que se equiparan con los descriptores o con el texto completo de los documentos.

En el caso de la Web el problema es básicamente el mismo (aunque agravado por las características propias de la Web) y la solución que en muchos casos se aplica es similar a la adoptada en las bibliotecas: el desarrollo de sistemas de filtrado de información (también conocidos como sistemas de recomendación). De hecho, los sistemas de recomendación podrían considerarse sistemas de DSI aplicados a la Web, aunque con capacidades de filtrado infinitamente más potentes y sofisticadas. Estos sistemas aplican técnicas de filtrado de información que facilitan el acceso de los usuarios a la información que necesitan. En dominios textuales, los sistemas de filtrado evalúan y criban los recursos disponibles en la Web (normalmente en formato HTML o XML) básicamente para asistir a los usuarios en tareas de recuperación de información (Resnick, Varian 1997) (principalmente mediante el uso de agentes de filtrado),

---

aunque también se utilizan para predecir la valoración de los usuarios sobre ítems que aún no han evaluado (Szomszor 2007).

Independientemente del dominio de aplicación, estos sistemas presentan una serie de características comunes (Callan et al. 2003):

- se aplican a información no estructurada o poco estructurada.
- se basan en los perfiles de usuario para adaptar la información a las necesidades de cada individuo.
- manejan grandes volúmenes de información.
- trabajan principalmente con información textual.
- su objetivo es eliminar información irrelevante de los resultados presentados a los usuarios.

El origen de este tipo de sistemas se remonta a principios de la década de los 90 del pasado siglo, cuando comienzan a surgir dentro de los servicios de *newsgroups* (grupos de noticias) servicios de filtrado de noticias que permitían a su comunidad de usuarios acceder exclusivamente a aquellas que potencialmente podían ser de su interés (Foltz, Dumais 1992) (Resnick et al. 1994) (Stodolsky 1990).

En la actualidad los sistemas de recomendación han evolucionado y es posible encontrarlos en diversos ámbitos de aplicación como en el comercio electrónico (donde se han convertido en una herramienta fundamental para los proveedores en línea -(Schafer, Konstan, Riedl 2001)-), o en los servicios de información.

No obstante, en cada dominio se presentan diferentes problemas a los que hay que dar diferentes soluciones. Esa capacidad de evolución ha provocado que los sistemas de recomendación se hayan diversificado y para dar soluciones diferentes a diferentes problemas, lo que ha dado lugar a diferentes tipos de sistemas de filtrado.

---



### 3.2.1 CRITERIOS DE CLASIFICACIÓN

Tradicionalmente los sistemas de filtrado y recomendación se han clasificado en tres categorías principales de acuerdo a la técnica que utilizan para realizar el filtrado (Popescul et al. 2001): sistemas de recomendación sociales, sistemas de recomendación basados en contenido y sistemas de recomendación basados en factores económicos. Veamos cada uno de ellos con mayor detalle.

Los sistemas de filtrado social, también denominados de filtrado en colaboración o *colaborativos* (anglicismo con el que más comúnmente son conocidos), utilizan la información proporcionada por usuarios de características similares que pertenecen a una misma comunidad para generar recomendaciones, obviando el contenido de los recursos (se basa exclusivamente en las valoraciones que estos reciben por parte de los usuarios). En este tipo de sistemas es habitual agrupar a los usuarios en categorías específicas o estereotipos que los caracterizan a través de una serie de valores de preferencia definidos por defecto y que representan las necesidades de información y hábitos de búsqueda más comunes del grupo. Este tipo de sistemas tiende a ofrecer resultados pobres cuando se dispone de poca información sobre los usuarios o estos tienen gustos muy heterogéneos (Popescul et al. 2001).

Los sistemas de filtrado basados en contenido generan recomendaciones equiparando las preferencias del usuario (expresadas por este de forma implícita o explícita) con los metadatos o características utilizados en la representación de los recursos o productos, ignorando de esta forma la información relativa a otros usuarios. Estos sistemas, al igual que los sociales, son poco fiables cuando se dispone de poca información sobre el usuario, por lo que es frecuente recurrir al uso de perfiles que caracterizan a los usuarios a través de sus preferencias explícitas o implícitas. De hecho, la integración de los perfiles de usuario en los procesos de filtrado supone una potente herramienta para proporcionar un filtrado rápido y eficiente (Shapira et al. 1997).

---

Por su parte, los sistemas de recomendación basados en factores económicos son aquellos que generan recomendaciones basándose en elementos de coste (Resnick et al. 1994), como, por ejemplo, la relación entre el coste de un servicio y el beneficio que reporta al cliente, o la relación entre el ancho de banda y el tamaño de un objeto a descargar. No obstante, el uso de este tipo de sistemas a día de hoy es muy marginal.

La tendencia actual se inclina por el desarrollo de sistemas de filtrado híbridos que combinan características de los sistemas basados en contenido y de los colaborativos para atenuar las desventajas de cada una de ellos y mejorar de esta manera la eficiencia global del funcionamiento del sistema en términos de precisión y exhaustividad (Basu, Hirsh, Cohen 1998) (Balabanovic, Shoham 1997).

No obstante, existen otros muchos criterios que consideran diferentes aspectos de este tipo de sistemas para clasificarlos.

Así, por ejemplo, atendiendo a la forma en que se capturan las preferencias del usuario, se distingue entre sistemas donde la recogida de datos es explícita (cuando se pide al usuario que aporte de forma voluntaria sus valoraciones) y sistemas donde esta recogida se realiza de manera implícita (monitorizando la actividad del usuario en el sistema).

Según la metodología de filtrado de información es posible distinguir entre sistemas que realizan un filtrado pasivo (Rafter, Bradley, Smyth 1999) (cuando se genera una única recomendación que es válida para todos los usuarios del sistema), y sistemas de filtrado activo (Boutillier, Zemer, Marlin 2003) (Maltz, Ehrlich 1995) en el que la recomendación se genera a partir del historial de recomendaciones de los usuarios para generar nuevas recomendaciones personalizadas. Un ejemplo de sistemas pasivos son aquellos que recomiendan los ítems más valorados por el conjunto de la comunidad de usuarios (es decir, el usuario no influye de forma directa sobre la información que recibe). Los sistemas activos son algo más complejos ya que generan las recomendaciones a partir de las opiniones de usuarios de perfil similar. Además dentro de los

---

sistemas activos es posible definir dos modelos diferentes de recuperación de información (Gnasa et al. 2005): el modelo de *pull* de información (cuando el usuario debe lanzar una consulta al sistema para recibir la recomendación), y el modelo de *push* de información (cuando esa consulta se hace de forma implícita a través de las preferencias definidas en su perfil de usuario).

También se suele distinguir entre sistemas de filtrado centrado en el usuario (Xin et al. 2005) cuando las recomendaciones equiparando la similitud entre usuarios de acuerdo a las preferencias almacenadas en su perfil, y sistemas de filtrado centrados en ítems o productos (Resnick et al. 2004), en los que primero se buscan las relaciones que existen entre los diferentes ítems y a continuación se generan las recomendaciones a partir de las preferencias del usuario activo.

Atendiendo a la manera en que el algoritmo de filtrado procesa la información encontramos sistemas que cargan todos los datos de forma residente en memoria para generar las recomendaciones, y otros que utilizan una cantidad de datos menor a partir de la cual se generan modelos predictivos que permiten procesar las recomendaciones de una manera más rápida.

Otro criterio consiste en diferenciar entre sistemas centralizados (cuando las descripciones de productos y perfiles de usuarios se encuentran almacenados en un servidor centralizado) y sistemas de recomendación no centralizados (generalmente desarrollados sobre redes P2P).

Como vemos, existen múltiples criterios de clasificación la mayoría de ellos no excluyentes entre sí permitiendo de esta manera el desarrollo de sistemas de filtrado híbridos.

No obstante, independientemente de la tipología estos sistemas adolecen de algunas carencias:

- i) los procesos de comunicación entre agentes, y entre agentes y usuarios se ven dificultados por las diversas formas en que la información está representada;
-

- ii) esta heterogeneidad en la representación de la información provoca a su vez que esta no pueda ser reutilizada en otros procesos y por otras aplicaciones;

Para solventar estas deficiencias es posible aplicar diferentes soluciones que facilitan los procesos de comunicación, como las técnicas de modelado lingüístico difuso con las que es posible definir un modelo de representación de la información flexible (fácil de interpretar por los usuarios y con la que pueden operar las máquinas) para procesar información cualitativa mediante el uso de etiquetas lingüísticas, o las tecnologías de Web Semántica que proporcionan los medios necesarios para definir recursos semánticamente accesibles tanto para las máquinas como para los consumidores de información.

Precisamente la irrupción de la Web Semántica ha dado lugar a la aparición en los últimos años de una tipología de sistemas de recomendación, los sistemas de recomendación semánticos, que proponen nuevas soluciones a sus problemas específicos de filtrado.

En el siguiente apartado pasamos a comentar sus principales características y las diferentes sub-tipologías que hemos detectado en la literatura (Peis, Morales-del-Castillo, Delgado-López 2008).

### **3.2.2 SISTEMAS DE RECOMENDACIÓN SEMÁNTICOS**

Este tipo de sistemas de filtrado y recomendación se caracterizan por basar su funcionamiento sobre una base de conocimiento, normalmente definida a través de un esquema de conceptos (como una taxonomía o un tesoro) o una ontología, y que utilizan tecnologías de Web Semántica.

El uso de esquemas de conceptos y ontologías en este tipo de sistemas permite paliar determinados problemas, entre los que se incluyen los siguientes (Carmagnola et al. 2006) (Middleton 2002):

---

- Garantizan la interoperabilidad de los recursos del sistema y la homogeneidad de la representación de la información.
- Permiten contextualizar de forma dinámica las preferencias de los usuarios en un dominio específico.
- Facilitan el trabajo en redes sociales y el filtrado colaborativo.
- Mejoran los procesos de comunicación entre agentes, y entre agentes y usuarios.
- Permite paliar el problema de “arranque el frío” al poder completar la información incompleta mediante inferencias.
- Permiten extender semánticamente las descripciones de los factores contextuales en que se encuentra el usuario.
- Mejoran la representación y descripción de los diferentes elementos del sistema.
- Mejoran la descripción de la lógica del sistema al admitir la inclusión de conjuntos de reglas (generalmente definidos en SWRL -(Horrocks et al. 2004)-).
- Proporcionan los medios necesarios para generar descripciones enriquecidas de servicios web y facilitar así a los agentes software su descubrimiento.

Hemos distinguido tres categorías básicas que a continuación procedemos a repasar: sistemas semánticos que ofrecen un único nivel de filtrado, sistemas adaptables al contexto y sistemas basados en redes de confianza.

#### **A) SISTEMAS DE RECOMENDACIÓN SEMÁNTICOS CON UN ÚNICO NIVEL DE FILTRADO**

Este tipo de sistemas de recomendación se caracterizan por proporcionar un único nivel de filtrado basado en el cálculo de medidas de similaridad de acuerdo a la ontología o esquema de conceptos sobre el que se apoya. La variedad de

---

este tipo de sistemas de recomendación semánticos es enorme y se aplican en múltiples dominios de aplicación. A continuación vamos a ver algunos ejemplos.

El sistema diseñado por Wang y Kong (2007) es un sistema de recomendación personalizado que intenta paliar los problemas de los sistemas de recomendación colaborativos usando, en forma de ontología, la información semántica de las características categóricas de un ítem. La similaridad por pares de usuario se calcula con el método de la media ponderada de tres medidas de similaridad: la similaridad del histórico de evaluaciones de dos usuarios (utilizando el coeficiente de correlación de Pearson sobre información de utilización del sistema en forma de matriz de datos de evaluación usuario-ítem); la similaridad de datos demográficos de dichos usuarios (calculada vía media ponderada); y, la similaridad del interés o preferencia de dichos usuarios basada en las similaridades semánticas de los ítems recuperados y/o evaluados. Al mismo tiempo, el sistema incorpora un mecanismo de *cluster* de usuarios *offline* para paliar el problema de la escalabilidad.

Khosravi, Farsani y Nematbakhsh (2006) sugieren una metodología para recomendaciones personalizadas en el contexto del comercio electrónico. Se trata de un procedimiento para recomendar productos a clientes potenciales. El algoritmo propuesto se basa en el modelado de información sobre productos y usuarios con OWL (Ontology Web Language). El proceso se inicia con la clasificación de productos y consumidores mediante OWL, lo que facilitará el análisis de la similaridad producto-cliente. En una segunda fase se seleccionan consumidores activos, teniendo en cuenta recomendaciones anteriores (el sistema no recomienda a un cliente si el número de sus recomendaciones anteriores no sobrepasa un cierto umbral). La clasificación de productos y clientes es utilizada para crear una matriz de evaluaciones productos-clientes. El algoritmo recomienda alguno de los productos de cada clase de entre las clases de productos basándose en el número de evaluaciones en la matriz.

---

Otro modelo que se aplica en el ámbito del comercio electrónico es el que presenta Ziegler, Lausen y Schmidt-Thieme (2004). El sistema se basa en el paradigma de recomendación colaborativa a través de contenido (Pazzani, 1999) y utiliza una taxonomía de productos a partir de la cual se definen los perfiles de los usuarios (sin necesidad de que estos proporcionen sus valoraciones de forma explícita). El perfil del usuario activo es utilizado para descubrir usuarios con intereses similares, cuyas valoraciones sirven al sistema para generar las recomendaciones.

Por su parte Jung y colaboradores (2005) proponen un sistema de recomendación basado en información personal que afirman es adecuado al contexto Web Semántica. El modelo se fundamenta en la representación de servicios Web y perfiles de usuario con tripletas RDF (Resource Description Framework). Cada compañía que quiera proporcionar servicios Web registra los datos en el repositorio de información, donde el sistema convierte los datos en documentos con formato RDF. El módulo de búsqueda extrae la información del repositorio y la envía al agente colector de documentos. El agente accede al espacio de nombre correspondiente y recoge los documentos RDF relativos a los servicios Web requeridos. Estos documentos son enviados al agente de integración de información, donde son fusionados en un solo documento RDF que contiene información relevante. Finalmente, el agente de recuperación de información extrae las tripletas RDF más relevantes conforme al perfil de usuario y se ofrecen al usuario los objetos almacenados coincidentes con dichas tripletas.

Otros sistemas se definen sobre estructuras descentralizadas como las redes P2P. Ese es el caso del modelo que presenta Díaz-Avilés (2005), donde la información no está disponible en un repositorio centralizado sino en cada uno de los componentes de la red. Los ítems u objetos se modelan mediante una ontología común que utilizan todos los miembros de la red. La selección de los componentes de la red se realiza de forma dinámica y la recomendaciones se generan utilizando un algoritmo de recomendación basado en el método del

---

“vecino más cercano” que se ejecuta de forma local en cada componente de la red.

Una aproximación original es la propuesta de Cantador y Castells (2006) para desarrollar un modelo de red social semántica multicapa que permite definir el sistema desde diferentes perspectivas a partir de los intereses comunes que comparten los integrantes de la red. A partir de una serie de perfiles de usuario generados utilizando una ontología de conceptos, y teniendo en cuenta sus preferencias comunes, el sistema es capaz de acotar los diferentes grupos de conceptos del dominio. A partir de estos grupos es posible identificar conjuntos de usuarios con intereses similares que se interrelacionan entre sí en diferentes niveles semánticos (de acuerdo a sus preferencias). Esta metodología permite descubrir redes sociales implícitas que pueden ser aprovechadas para definir tanto sistemas de recomendación basados en contenido, como colaborativos.

Frente a estos modelos genéricos, encontramos muchos otros que se definen para dominios más específicos como el sistema que proponen Middleton y colaboradores (2002), para recomendar artículos científicos, que se basa en una ontología generada de forma automática a partir de la información extraída dinámicamente de diversas fuentes (información en línea, monitorización de usuarios, feedback, etc.); el sistema multi-agente *AVATAR* (Blanco Fernández et al. 2004) especializado en programas de televisión; o *SemMF* (Oldakowsky, Byzer 2005) que genera recomendaciones de ofertas de trabajo calculando la similaridad semántica de dos conceptos de acuerdo a su situación dentro de una jerarquía de conceptos.

## **B) SISTEMAS DE RECOMENDACIÓN SEMÁNTICOS Y LA WEB 2.0**

El éxito de plataformas como la Web 2.0 donde interaccionan personas que establecen entre sí redes sociales y comparten información que les es de interés, ha hecho que los sistemas de recomendación (sobre todo los colaborativos) se

---



hayan convertido en una herramienta de gran utilidad. De igual manera, los sistemas de recomendación semánticos se están empezando a implantar con bastante éxito, ya que las ontologías y esquemas de conceptos permiten describir el conocimiento común que comparten las diferentes comunidades de usuarios que conforman la base de la Web 2.0

Así por ejemplo, Adrian, Sauermann y Roth-Berghofer (2007) proponen *ConTag*, un sistema que asesora a los usuarios sobre las etiquetas más propicias para describir el contenido de diferentes recursos que se utilizan en una plataforma Web 2.0. Los documentos son transformados a formato RDF para posteriormente generar un mapa de materias (*topic map*) definido en SKOS Core, a partir del cual es posible calcular la similaridad de estas materias con los conceptos definidos en la ontología del sistema. Las etiquetas recomendadas por el sistema pueden ayudar al usuario a describir mejor el contenido del recurso, minimizando así problemas de sinonimia, homonimia, acrónimos y variantes ortográficas que pueden presentar las etiquetas definidas por los usuarios. Jäschke y colaboradores (2007) proponen un mecanismo de recomendación de etiquetas similar.

Szomszor y colaboradores (2007) proponen el uso de algoritmos de predicción para la integración de una *folksonomía* de películas con una base de conocimiento semántico sobre alquileres de películas por parte de los usuarios. La *folksonomía* es usada para enriquecer la base de conocimiento con descripciones y categorizaciones de títulos de películas, además de con representaciones de opiniones e intereses de los usuarios. La “nube” de etiquetas generada por la *folksonomía* es usada para construir perfiles de usuario mejorados que reflejan el nivel de interés del usuario sobre diferentes tipos de películas y, por lo tanto, proporcionan una base para la predicción de calificaciones de películas no vistas. Para la construcción de la base de conocimiento sobre películas y sobre cómo las alquilan los usuarios emplean fuentes de la Web 2.0 (*Internet Movie Database* y *Netflix Prize*,

---

respectivamente). Para proporcionar una representación homogénea de ambos conjuntos de datos se construye una ontología con OWL, en conjunción con la tecnología de “*mapeado*” D2RQ, que permite tratar los datos de una base de datos relacional como grafos virtuales RDF.

En la literatura también encontramos otros modelos de sistemas de recomendación que utilizan tecnologías propias de la Web 2.0 como los canales RSS. Como el modelo presentado por Kobayashi y Saito (2006) que permite recomendar artículos periodísticos y noticias extraídos de canales RSS, utilizando como apoyo un tesoro cuyos términos se usan para representar tanto las materias de las noticias como las necesidades de información de los usuarios, o el proyecto «*Foafing the Music*» (Celma, Ramirez, Herrera 2005) que recomienda temas musicales y noticias relacionadas que se extraen de canales RSS y otras fuentes de información de la Web.

### **C) SISTEMAS BASADOS EN REDES DE CONFIANZA**

Como hemos visto, una de las principales preocupaciones de los investigadores en el campo de los sistemas de recomendación semánticos es garantizar la fiabilidad y precisión de las recomendaciones que se generan. Muchos de estos sistemas ofrecen un nivel de filtrado adicional en el que partiendo de las preferencias definidas en el perfil del usuario (generado a partir de una ontología de conceptos) se crean redes de confianza en la que los agentes de recomendación deciden, de acuerdo al nivel de confianza que les merece el resto usuarios, si las recomendaciones que estos pueden ofrecer son fiables y de calidad (Ziegler 2004).

En esta línea Massa y Avesani (2004) presentan un modelo basado en la denominada “Web de Confianza”. El sistema toma como inputs una “matriz de confianza” (que representa el conjunto de declaraciones de confianza de la comunidad –los usuarios explicitan a aquellos otros en cuyas evaluaciones

---

confían-) y una matriz de evaluaciones (que representa todas las evaluaciones dadas por los usuarios a los ítems) y produce, como output, una matriz de evaluaciones predecibles que los usuarios podrían asignar a los ítems. Esta matriz es usada por el sistema para recomendar los ítems preferidos por los usuarios. Es decir, el sistema selecciona de la columna de evaluaciones predecibles relativa a los usuarios, los ítems con más altos valores. De esta forma, se incrementa la cobertura de los sistemas de recomendación colaborativos y, al mismo tiempo, se mantiene la calidad de los ítems recomendados, paliando así, los problemas de arranque en frío (para nuevos usuarios) y la nula fiabilidad de los datos de usuarios maliciosos.

Una aproximación diferente es la que ofrece el sistema FOAFRealm (Kruk, Decker 2005). Se construye sobre redes P2P y que dispone de una librería que permite a los usuarios gestionar sus propios perfiles definidos en FOAF. A su vez, las descripciones y categorías definidas por los usuarios están enriquecidas con términos extraídos de esquemas taxonómicos u ontologías. Para gestionar los problemas de seguridad que suelen presentarse en los sistemas de recomendación basados en arquitecturas distribuidas se define el concepto de “perfil extrapolado”, en el que los nuevos usuarios adquieren un perfil que se construye a partir del perfil de los usuarios que lo agregan como “amigo”, y que depende directamente del nivel de confianza que estos le otorgan (facilitándose de esta manera la localización de usuarios maliciosos).

De manera similar funciona Filmtrust (Golbek 2005), un sitio web que integra un sistema de recomendación semántico colaborativo de películas que utiliza el vocabulario FOAF como base para la creación de una red social de confianza. En esta red cada individuo debe valorar la confianza que les merecen aquellos usuarios a los que añaden a su red de amigos (a partir de los cuales se generan las recomendaciones de películas). El sistema adicionalmente dispone de un mecanismo que permite determinar la precisión de las recomendaciones generadas.

---

Bedi y colaboradores (2007) definen un modelo novedoso que se basa en el uso de una serie de ontologías y del establecimiento de redes de confianza entre agentes. Cada agente tiene asociada una ontología personal temporal (de esta manera se amortiguan las posibles modificaciones que estas pueden sufrir en el tiempo) a partir de las cuales pueden generar recomendaciones independientes de dominio. Para calcular estos valores de recomendación la similaridad entre agentes se presupone implícitamente en el valor de confianza que los propios agentes se asignan entre sí al interactuar. También es novedosa la utilización en este tipo de sistemas de conjuntos difusos intuitivos (Atanassov, 1999) para gestionar la incertidumbre inherente al proceso de recomendación.

#### **D) SISTEMAS ADAPTABLES AL CONTEXTO**

Otra gran categoría de sistemas que se puede establecer dentro de los sistemas de recomendación semánticos son los sistemas de recomendación adaptables (o sensibles) al contexto. En este tipo de sistemas se analizan y toman en consideración diferentes factores (temporales, de lugar, nivel de experiencia del usuario, dispositivo que se está utilizando en el momento de recibir la recomendación, etc.) para inferir el contexto en que se encuentra el usuario y adaptar las recomendación a esas circunstancias.

Dentro de este grupo encontramos, por ejemplo, el modelo basado en contenido que proponen Kim y Kwon (2007). El funcionamiento del sistema se basa en la definición de “contextos de uso” que se corresponden con los diferentes niveles de especificidad de una ontología de conceptos. El sistema genera una recomendación a partir del conjunto de ítems más valorados por un usuario y a continuación es capaz de adaptar el nivel de especificidad de la información que es mostrada al individuo dependiendo del contexto de uso en que se encuentran los conceptos por los que muestra interés. Para determinar estos contextos el sistema utiliza cuatro ontologías: una ontología de productos, otra donde se

---

definen los diferentes contextos de uso, una tercera sobre el registro histórico de actividades de los usuarios en el sistema, y una última ontología sobre los usuarios.

Loizou y Dasmahapatra (2006) proponen por su parte un sistema fundamentado sobre una ontología que incorpora información contextual tanto del proceso de recomendación como de los ítems a recomendar. Esta información contextual (por ejemplo, el tiempo de recomendación o la utilidad del ítem recomendado para el usuario) tratada con un aparato matemático basado en reglas heurísticas aplicadas sobre espacios vectoriales permite al sistema evaluar, de forma dinámica, la adecuación de una recomendación concreta.

También es interesante la propuesta de Yu y colaboradores (2007) que aplican este tipo de sistemas al ámbito del e-learning. El modelo pretende facilitar a los estudiantes los recursos que necesitan para llevar a cabo las tareas planteadas en los cursos. Se apoya en el uso de ontologías para representar de manera eficiente el conocimiento del sistema sobre los usuarios, sobre el contenido de los recursos y sobre el dominio específico de especialización del sistema para generar recomendaciones basadas en el contexto de uso (es decir, la recomendación es diferente dependiendo de diferentes factores como el nivel de experiencia, o de lo avanzado que esté el curso en el momento en que el usuario requiere la recomendación). El sistema no se limita a recomendar recursos adecuados al contexto, sino que sugiere al alumno otros recursos relacionados que le permiten ahondar en la materia y definir de esta manera un programa de estudios completo.

Otros modelos que aplican este tipo de filtrado son el propuesto por Laliwala, Sorathia y Chaudhary (2006) donde se desarrolla un sistema de recomendaciones semántico para servicios de información agrícola basado en eventos o el que presentan Woerndl, Schueller y Wojtech (2007) donde las ontologías son utilizadas para mejorar la descripción de servicios Web ofrecidos

---

por terceros, y para desarrollar un servicio de recomendaciones colaborativo especializado en información turística para dispositivos móviles, capaz de gestionar tanto información estática (definida en los perfiles de los usuarios) como información dinámica (contextual).

#### **E) MEDIDAS DE SEMEJANZA SEMÁNTICA**

Un aspecto que va íntimamente ligado al desarrollo de los sistemas de filtrado y recomendación son las medidas de semejanza o similitud. Dado que la web Semántica introduce un nuevo paradigma de representación de la información las medidas de similitud tradicionales pierden su eficiencia y es necesario buscar nuevas formas de comparar objetos informativos (como recursos o perfiles de usuario).

Podríamos decir que el filtrado consiste básicamente en la equiparación de una serie de características presentes, tanto en el perfil del usuario como en los recursos de los que el sistema dispone, cuya finalidad es ofrecer al individuo aquellos recursos que más se asemejan a sus preferencias. El modelo más extendido para realizar el filtrado es el modelo vectorial, en el que perfiles y documentos se representan como vectores de términos cuya distancia en el espacio vectorial define su grado de similitud. Para calcular la semejanza entre ambos existen diferentes medidas de similitud, como la del Coseno (Salton 1971) (Salton, Wong, Yang 1975), Dice (van Rijsbergen 1979) y Jaccard (Jaccard 1912) (Rorvig 1999).

En el marco que propone la Web Semántica, donde se utilizan estructuras de organización del conocimiento, como las ontologías y las taxonomías, han aparecido diferentes aproximaciones al problema del cálculo de la distancia, ya no entre vectores de términos, sino entre vectores de conceptos (teniendo en cuenta su posición en estas estructuras jerárquicas), entre ontologías (Maedche,

---

Staab 2002) (Ehrig, Haase, Stojanovic 2004) o entre servicios web semánticos (Hau, Lee, Darlington 2005).

Si nos centramos en las medidas de semejanza entre conceptos encontramos trabajos como el de Savia, Kurki y Jokela (1998), donde para calcular la distancia entre conceptos en una taxonomía, a cada uno de ellos se les asigna un peso que es proporcional al número de términos específicos que de ellos dependen. En esta misma línea, Ziegler (2004) propone que cada concepto tiene un peso que depende de su especificidad o profundidad dentro del árbol jerárquico. Otros métodos basan la medida de la similitud de conceptos en el contenido informativo de estos. Resnick (1999) asocia a cada concepto la probabilidad de encontrar una instancia suya en la taxonomía. La probabilidad está afectada por un logaritmo negativo, de forma que cuanto mayor sea ésta (es decir, cuanto más genérico o abstracto sea un concepto) menor será su contenido informativo. Existen también otros métodos, como el propuesto por Oldakowsky y Byzer (2005) (que es el que utilizamos en el modelo que proponemos en el capítulo 4), en el que la distancia entre conceptos viene representada por el camino que hay que seguir para llegar desde el concepto *a* al concepto *b* a través de un antecesor jerárquico común en una taxonomía.

El desarrollo de estas técnicas y modelos suponen un avance cualitativo que permite desarrollar servicios de información sobre infraestructuras semánticas.

### **3.3 SINDICACIÓN DE CONTENIDOS**

Ante la aparición de nuevas necesidades por parte de los consumidores de información, los proveedores se ven en la tesitura de tener que ofrecer servicios de valor añadido que satisfagan estos requerimientos de una manera más eficiente. Un ejemplo de este tipo de servicios son servicios de redifusión o *sindicación* de contenidos.

---

La redifusión de contenidos consiste en facilitar a los usuarios el acceso a los contenidos de páginas web que se actualizan con frecuencia. Supone una opción para navegar por múltiples sitios web, consultando información actualizada, sin la necesidad de visitar todos y cada uno de esos sitios.

Este tipo de servicios tienen su origen en las tecnologías de *push* (o recepción pasiva) de información en las que un servidor distribuye contenidos de acuerdo a unos parámetros definidos en los clientes sin que sea necesaria una petición explícita por parte del usuario para que el servidor mande dicha información.

Esto se consigue mediante la creación de canales o *feeds* que contienen una serie de ítems que describen los diferentes contenidos que se quieren difundir y que enlazan a dichos contenidos. Para generar estos canales en la actualidad existen varios vocabularios que permiten la gestión ágil y flexible de listados de hiperenlaces.

Aunque existen precedentes de varios lenguajes específicos para la redifusión de contenidos, como Channel Definition Format [137] o ScriptingNews (Winner 1999), hubo que esperar hasta 1999 para que Dan Libby creara RSS 0.9 (*Rich Site Summary*) (Libby 1999) para el portal MyNetscape.com. Las versiones 0.91 [138] y 0.92 [139] han sido desarrolladas por el equipo UserLand, liderado por Dave Winner, que tras abandonarlo diseñó en solitario la especificación RSS 2.0 [140]. Caso aparte es el del equipo independiente RSS-Dev que desarrolló la especificación RSS 1.0 basándose en el modelo de datos RDF y alejándose de esta forma de la filosofía de desarrollo centrada exclusivamente en XML.

A pesar de que coexisten diferentes versiones de RSS (y otros formatos como Atom (Nottingham 2005), los dos únicos vocabularios que mantienen un desarrollo estable son RSS 1.0 (RDF Site Summary) y RSS 2.0 (Really Simple Syndication). Ambos utilizan la sintaxis XML y disponen de módulos que permiten extender su vocabulario sin modificar el núcleo de elementos. No obstante, la principal diferencia entre ellos radica en que RSS 1.0 utiliza el

---



modelo de datos de RDF para estructurar la información (lo cual permite hacer los canales RSS 1.0 semánticamente accesibles).

Hoy día es muy común encontrar servicios de redifusión de contenidos tanto textuales como multimedia (también conocidos como *Podcasts*) ya no solo en blogs personales, sino también en grandes portales de información (la mayor parte de revistas, periódicos, radios, televisiones y sitios web de noticias disponen de canales RSS que se actualizan periódicamente con noticias de actualidad), en instituciones y organismos públicos (como los del Ministerio de Cultura [141]), en instituciones de educación como centros de e-learning o universidades [142], en sitios web comerciales (como eBay [143]), o en centros de información como las bibliotecas donde, entre otros usos, los canales RSS se utilizan como herramientas para la difusión de servicios y adquisiciones o como servicios de DSI basados en perfiles estáticos (Sánchez 2007). También en dominios como el académico-científico los servicios de redifusión de contenidos pueden ser de especial utilidad, ya que permiten mantener informados a los investigadores de una manera inmediata con contenidos muy actualizados extraídos de diferentes fuentes (especializadas como los canales de publicaciones reputadas o más informales como foros o blogs de expertos).

### **3.3.1 CARACTERÍSTICAS**

Si existe una característica a destacar de este tipo de servicios ese es sin duda la sencillez ya que:

- Acceder a los canales RSS es sencillo. El único requisito que debe cumplir el consumidor de información para recibir contenidos es suscribirse previamente a un canal de su interés. Este proceso resulta muy simple ya que basta con visitar el sitio web de cuyos contenidos desea mantenerse informado o bien usar alguno de los múltiples servicios de búsqueda o directorios que permiten descubrir los canales RSS, respectivamente, utilizando palabras clave, o bien navegando directamente por una
-

estructura jerárquica de categorías temáticas). Algunos de estos servicios (muchos de ellos híbridos, ya que ofrecen ambas alternativas de búsqueda) son RSS Network [144], RSS News Feed Directory [145], Search4RSS [146], o Syndic8 [147]).

- La publicación de canales RSS resulta un proceso muy sencillo. Basta con generar el código del canal en cualquiera de los vocabularios y alojarlo en un sitio web. No obstante, suele resultar más eficaz (rápido y económico) difundir y hacer visibles los canales utilizando los denominados sistemas de *ping* que permiten alertar a uno (o varios) directorios de canales RSS de que los contenidos de un sitio web han sido actualizados. De esta manera forzamos a los directorios a indizar nuestro canal sin tener que esperar a que lo haga su araña web. Ejemplos de estos sistemas son Autopinger [148] o Pingomatic [149]
- La visualización de los canales RSS también resulta muy sencilla. Existen multitud de programas capaces de procesar y presentarlos en pantalla. Sánchez (2007) clasifica estos programas en seis categorías:
  - Lectores
  - Agregadores
  - Plug-ins para navegadores
  - Navegadores
  - Servicios web
  - Sistemas de gestión de contenidos con módulos lectores de RSS.

De hecho, una de las claves del éxito de este tipo de servicios radica en que la mayoría de los navegadores web ya poseen lectores RSS integrados, por lo que ni siquiera es necesario recurrir a software adicional para acceder a ellos.

Otra característica interesante de los servicios de redifusión de contenidos es la posibilidad de compartir con otros usuarios el listado de canales RSS a los que un individuo está suscrito mediante el intercambio de ficheros OPML (Online

---

Processor Markup Language) [150]. Estos ficheros también pueden ser interpretados directamente por agregadores y lectores RSS.

### **3.3.2 FILTRADO DE CANALES RSS**

Sin embargo, y a pesar de sus grandes ventajas, el principal inconveniente de los servicios de redifusión de contenidos es una vez más la sobrecarga de información ya que, aunque la suscripción a un canal de una temática determinada se puede considerar como un primer nivel de filtrado, es complicado controlar la idoneidad del contenido de todos los ítems que recibe un suscriptor. Los servicios de filtrado RSS ofrecen una solución parcial a este problema.

Muchos de estos servicios se ofrecen en línea por lo que es necesario indicar la URL del canal a filtrar. Entre ellos encontramos algunos que realizan el filtrado mediante la introducción de palabras clave simples definidas por el usuario que se buscan a texto completo en el canal, como es el caso de ReFilter [151], o bien permitiendo definir consultas algo más complejas con operadores booleanos seleccionados de forma implícita en la formulación de dichas consultas (como ocurre con Feed Sifter [152]). Otras aplicaciones refinan algo más el filtrado, como FilterMyRSS [153], permitiendo definir el área informativa de los ítems del canal donde debe ser buscada la palabra (en el título, el resumen o el área de materias).

Otro tipo de servicios se limitan a filtrar exclusivamente los contenidos de una serie de canales por defecto (por lo general de temática variada) que el sitio web pone a disposición a sus usuarios. En esta categoría encontramos servicios como Pulverati [154] o Feed Collectors [155], donde la colección de canales está formada por canales creados por la propia comunidad de usuarios del sitio web.

---

Existen otro tipo de servicios, como ZAPTXT [156], que combinan varias de estas funcionalidades y permiten realizar búsquedas sobre grupos de canales (recomendados por el sitio web, definidos por el usuario o contenidos en un fichero OPML) utilizando categorías, etiquetas, palabras clave y operadores booleanos para combinarlas. Como servicio adicional, cuando se produce alguna actualización, el usuario es alertado a través de un mensaje SMS, mediante mensajería instantánea o a través de su correo electrónico.

Existen otros servicios que enfocan el filtrado desde otra perspectiva. Por ejemplo, AideRSS [157], utiliza una metodología diferente para filtrar los canales ya que esta se realiza de acuerdo al grado de interés que un determinado tema suscita en el usuario. Sobre un tema específico (y en un canal determinado) el usuario puede decidir disponer de todos los artículos (ideal para usuarios que desean una gran cobertura del área temática), recibir solo los calificados como “*buenos*” (tienen un grado de relevancia medio y alto, y por lo tanto son los adecuados para el perfil de usuarios con un alto nivel de interés sobre el tema), los artículos calificados como “*muy buenos*” (aquellos con un grado de relevancia mediano, idóneos para usuarios con un nivel medio de interés en el tema), o bien solo aquellos artículos valorados como “*los mejores*” (artículos con un alto grado de relevancia adecuado cuando el nivel de interés del usuario es solo puntual). Evidentemente, el nivel de exhaustividad (y por lo tanto de artículos mostrados) decrece con el nivel de interés. De manera opcional, si el usuario solo desea ver los artículos más valorados dispone de la opción de visualizar exclusivamente los 20 artículos más relevantes del canal.

También son populares las herramientas que permiten generar nuevos canales RSS a partir de la agregación de canales extraídos de diferentes fuentes (referencias URL o ficheros OPML), sobre los que previamente se ha realizado un proceso de filtrado utilizando palabras clave (Feed Findings [158], RSSFilter [159], MySyndicaat [160], Turtilla [161], Advanced RSS Mixer [162], *Blastfeed* [163]), o

---

aplicaciones que a esta funcionalidad añaden la posibilidad de ordenar y traducir los ítems (Yahoo! Pipes [164]).

También existen herramientas en línea que, como Feedbite [165], permiten crear filtros para eliminar anuncios o artículos no deseados de un canal RSS determinado.

Frente a los servicios web, encontramos otras aplicaciones que deben ser instaladas localmente y que por lo general disponen de herramientas potentes para filtrar el contenido de canales RSS. Así por ejemplo, encontramos RSS2HTML Scout [166] que permite el filtrado de los canales mediante el uso de palabras clave que aparezcan en el título, el enlace o en la descripción de los mismos, o Feed Rinse [167] que dispone de herramientas para definir reglas y bloquear artículos con características específicas (aquellos que contienen determinadas palabras o fueron publicados en una fecha específica), eliminar palabras malsonantes, y filtrar artículos mediante el uso de palabras clave o etiquetas, o según su autor.

Como vemos, la mayoría de estos servicios se limitan a ofrecer filtrado de información de acuerdo a una serie de parámetros que se definen en el mismo momento de la recuperación de ítems, o bien de acuerdo a unas preferencias definidas en perfiles estáticos, pero no hemos detectado ninguno que permita actualizar de forma dinámica las preferencias del individuo (uno de los pilares sobre los que se apoya el modelo que presentamos en el capítulo 4).

---

### **3.4 HERRAMIENTAS PARA EL DESARROLLO DE SERVICIOS DE VALOR AÑADIDO**

En este apartado pasamos a analizar algunas de las herramientas que sirven para el desarrollo de estos servicios de valor añadido para bibliotecas digitales, y que nos van a servir de base para el desarrollo del modelo y el prototipo que presentamos en esta memoria.

#### **3.4.1 LOS SISTEMAS DE ORGANIZACIÓN DEL CONOCIMIENTO**

La forma en que los recursos de información disponibles en la red están descritos o indizados es muy heterogénea, pudiendo encontrar diferentes vocabularios (algunos de ellos controlados, otros no) para definir materias incluso dentro de un mismo dominio. No solamente es heterogénea su representación, sino que también la forma en que estos recursos están organizados puede responder a diferentes esquemas de clasificación.

Los sistemas de organización y representación del conocimiento, más conocidos como KOS (*Knowledge Organization Systems*) son herramientas que permiten categorizar recursos de acuerdo a un esquema de organización determinado para facilitar su posterior recuperación, y que, de acuerdo a la definición de Zeng y Chan (2004), permiten al usuario interpretar estructuras de conocimiento de una manera organizada. Por lo tanto, además de una herramienta que permite mejorar la eficiencia en el funcionamiento de determinados servicios en un sistema de información, la mera presencia de este tipo de estructuras ya supone en sí misma un valor añadido para el usuario.

El interés que se ha despertado en los últimos tiempos alrededor de los KOS responde a la concurrencia de varios factores (Gilchrist 2003):

El exceso de información ha creado la necesidad de ofrecer al consumidor de información herramientas auxiliares que complementen a los buscadores web en la recuperación de información.

---

La falta de alfabetización informacional en el caso de muchos usuarios hace necesario el desarrollo de KOS como herramientas que faciliten a los usuarios la tarea de recuperar información.

La necesidad de disponer de herramientas que recojan la terminología específica que se utiliza en determinadas instituciones donde gran parte de la información que se consume es generada de forma interna.

Fruto de las fusiones y adquisiciones ha surgido en el ámbito empresarial el fenómeno del trabajo cooperativo a través de extranets y de las comunidades de trabajo virtuales. Muchos de estos grupos de trabajo son completamente ajenos a la cultura corporativa de la institución, por lo que compartir información y conocimiento entre grupos de usuarios tan heterogéneos se convierte en una tarea compleja.

Aunque el origen de la mayoría de los KOS se circunscribe al ámbito de las bibliotecas (y otros sistemas de información) tradicionales, la mayoría de ellos han sufrido un proceso de “migración” a la Web, formando parte del conjunto de herramientas disponibles, por ejemplo, en bases de datos de carácter bibliográfico o en bibliotecas digitales.

De acuerdo a características como la estructura y su complejidad, al tipo de relaciones que establecen entre sí los términos, y su función histórica los KOS se pueden clasificar en tres grupos básicos (Hodge 2000):

- Listas de términos: Contienen una enumeración de palabras, en ocasiones acompañadas de su definición correspondiente. Entre ellas encontramos los ficheros de autoridades, los glosarios, los diccionarios y los índices toponímicos.
    - Las listas de autoridades permiten controlar las diferentes variantes que permiten denominar una entidad (por ejemplo, países, o personas físicas o jurídicas) o un dominio específico de un área de conocimiento determinado. Los términos se suelen
-

presentar ordenados alfabéticamente o estructurados según un esquema de clasificación simple. No obstante, sí que disponen de mecanismos rudimentarios de navegación basados en una estructura jerárquica simple, y en las referencias que se definen entre los términos admitidos y los no admitidos.

- Los glosarios son listas de términos de un dominio determinado que suelen ir acompañados de definiciones o comentarios específicos para ese dominio.
  - Los diccionarios son herramientas con un alcance más general que los glosarios. Son listados alfabéticos de términos en los que se incluye su definición, sus variantes semánticas, y en ocasiones información sobre el origen del término, sus sinónimos o sus variantes ortográficas. No obstante estos términos relacionados no muestran ningún tipo de estructura jerárquica explícita.
  - Índices toponímicos: Son diccionarios que incluyen términos relativos a lugares geográficos y suelen aparecer como índices de atlas. Las entradas suelen estar identificadas por una categoría que las especifica (ciudad, río, cordillera, etc.) y proporcionan las coordenadas geográficas del lugar. A menudo se estructuran utilizando esquemas de clasificación por materias.
  - Clasificaciones y esquemas de categorización: Engloba aquellas herramientas que estructuran los términos en conjuntos temáticos, como los encabezamientos de materia, las taxonomías y los esquemas de clasificación y categorización.
    - Los encabezamientos de materia definen un sistema de términos controlados de amplia cobertura temática, organizados de acuerdo a una estructura jerárquica muy básica, que permite representar las materias de un recurso. Disponen de reglas para
-



combinar términos y de esta forma definir conceptos más específicos. Son un tipo de lenguaje de indización precoordinado, por lo que los términos que elijamos para indizar un recurso determinado se combinan en el mismo momento de la indización en un orden previamente establecido y la recuperación se realizará secuencialmente, siguiendo ese orden. De esta forma un usuario no podrá recuperar documentos a partir de aspectos que no estén definidos explícitamente en los propios encabezamientos.

- Taxonomías y esquemas de clasificación y categorización son términos que se suelen usar indistintamente para denominar a este tipo de esquemas, aunque existen sutiles diferencias entre ellos. El término “taxonomía” se utiliza en determinados ámbitos para referirse a agrupaciones de términos, objetos o seres vivos ordenados de forma jerárquica de acuerdo a una característica particular. Bajo esta denominación podemos encontrar, desde directorios web, a taxonomías generadas mediante procesos de categorización automática de documentos, o grandes taxonomías corporativas. Por su parte los esquemas de clasificación y categorización son términos utilizados para referirse a otros lenguajes de indización precoordinados como la Clasificación Decimal Universal (CDU) o la Clasificación decimal de Dewey. No obstante, los tres esquemas coinciden en definir categorías temáticas generales bajo las cuales se agrupan una serie de términos ordenados numérica o alfabéticamente, aunque no dispongan de una estructura jerárquica bien definida como en los tesauros. De hecho en algunos tesauros se suelen usar estos esquemas de clasificación para agrupar los términos del tesoro por categorías, y utilizarlos de esta forma como herramienta auxiliar de navegación.
-

- Listas relacionadas: Bajo esta denominación de encontramos herramientas que trabajan con listados de conceptos en los que se define para cada uno de ellos las relaciones y conexiones que establecen entre sí diferentes términos asociados a dichos conceptos. En este grupo de herramientas encontramos los tesauros, las ontologías y las redes semánticas.
    - Los tesauros son listas jerárquicas de términos o conceptos controlados pertenecientes a un dominio de conocimiento específico que se organizan de acuerdo a una estructura jerárquica, y que establecen relaciones semánticas entre sí. Con el control terminológico se pretende neutralizar la sinonimia y la polisemia características del lenguaje natural, que dificultan desarrollar de una manera precisa la indización y recuperación de información, y de esta manera facilitar al usuario la expresión escrita de ideas (Roget 1911). Para el usuario, el tesoro supone una valiosa herramienta que le asiste en el uso correcto de los términos necesarios para formular adecuadamente, extender o precisar sus consultas (Foskett 1997). De esta manera se puede minimizar el ruido documental, se pueden identificar los términos de indización con un sentido semántico claro al estar convenientemente contextualizados, y se abren las puertas a la recuperación mediante conceptos (lo cual representa una gran ventaja para los dominios de conocimiento verticales). Cuando un usuario quiere recuperar documentos relativos a un tema de su interés, crea una conceptualización de aquello que busca (es decir, genera una representación de su necesidad de información) que traslada a un lenguaje de recuperación. La misión del tesoro en este contexto es la de expandir la consulta sugiriendo términos relacionados que permitan enriquecer y precisar la búsqueda del usuario. Al contrario que las listas de autoridades, en el momento
-

de la recuperación de información la combinación de términos es postcoordinada, es decir, todos los conceptos que se extraen en la indexación para expresar la materia de un recurso tienen la misma categoría y no existen reglas para expresarlos en un orden determinado. Serán los propios usuarios los que definan la combinación de conceptos utilizando operadores booleanos o de comparación. El uso de la postcoordinación proporciona una mayor flexibilidad al usuario para definir sus estrategias de búsqueda que la precoordinación, aunque puede repercutir negativamente en la precisión de los resultados debido a que la naturaleza de la relación que une dos o más términos no se establece de manera expresa.

- Según la RAE [2], en su origen el término “ontología” denomina una rama de la filosofía que se encarga del estudio del ser en general, de sus propiedades trascendentales y, más específicamente, de la organización de la realidad. Como derivación de esta última acepción encontramos la definición de las ontologías entendidas como KOS. Según diferentes autores (Neches 1991) (Gruber 1995) (Guarino 1998) las ontologías pueden ser entendidas como la suma de una serie de conceptos relevantes que representan el conocimiento compartido por los miembros de un dominio concreto, las relaciones que establecen entre sí estos conceptos, y los axiomas definidos sobre estos conceptos y relaciones. De una manera más simple, podríamos definir ontología como la especificación explícita de una conceptualización o parcela de realidad. A pesar del paralelismo que por su definición se podría establecer con tesauros o redes semánticas, las ontologías suponen un salto cualitativo sobre los KOS tradicionales. Comparten con tesauros y redes semánticas el trabajar con conceptos relacionados, pero, no obstante, existe
-

una diferencia fundamental, y es que permiten definir relaciones semánticas complejas, reglas y axiomas que no están presentes en el resto de KOS.

- Las redes semánticas, a diferencia de tesauros y ontologías, no definen la estructura de los conceptos mediante jerarquías sino que utilizan redes formadas por nodos o vértices que simbolizan a los conceptos, y arcos o líneas que los unen para representar propiedades o relaciones entre conceptos. Estas redes se representan gráficamente mediante nodos orientados, en los que existe un nodo inicial y un nodo final de manera que la relación que existe entre los conceptos se define en una dirección determinada.

A continuación, vamos a analizar más profundamente la estructura y elementos de tesauros y ontologías, ya que son sistemas de organización del conocimiento que juegan un papel clave dentro del modelo de Web Semántica al ser las principales herramientas utilizadas para representar dominios de conocimiento.

#### **3.4.1.1 TESAUROS**

Según la definición de Van Slype (1991) *“un tesauro es una lista estructurada de conceptos destinados a representar de manera unívoca el contenido de los documentos y de las consultas dentro de un sistema documental determinado, y a ayudar al usuario a la indización de los documentos y de las consultas”*. En ella ya se definen claramente cuáles son los principales objetivos que debe cumplir el tesauro dentro de un sistema de información. El tesauro, al igual que otros lenguajes documentales, es un instrumento de control de vocabulario utilizado en un determinado sistema documental que sirve de puente entre el analista (que se encarga de describir los recursos disponibles en el sistema) y el usuario (que es el destinatario final de los recursos recuperados). En este contexto, su

---

objetivo primordial es eliminar la ambigüedad del lenguaje en la disciplina o el área temática a la cual pertenecen los términos.

Es posible distinguir los siguientes elementos estructurales en un tesoro (van Slype 1991):

a) *Unidades léxicas*

Se corresponden con las palabras que representan los conceptos que forman el tesoro. Se distinguen varias categorías:

- **Descriptores:** Son palabras o expresiones que representan de manera unívoca conceptos susceptibles de ser utilizados para indizar un documento o generar una consulta dentro del sistema documental. Por lo general son palabras aisladas (principalmente sustantivos), términos compuestos o frases.
  - **No descriptores:** Son términos tomados de una lista de sinónimos o cuasi-sinónimos de los descriptores (con los que establecen una relación de equivalencia semántica) o términos que designan en lenguaje común conceptos afines a los que definen los descriptores. Los no descriptores no pueden ser utilizados para la indización del documento o de las preguntas, pero reenvían a sus descriptores correspondientes.
  - **Descriptores auxiliares:** Son cualificadores que contextualizan la semántica del descriptor. Su misión original es evitar tener que definir un nuevo descriptor y reducir de esta manera el tamaño del tesoro.
  - **Grupos de descriptores:** Son conjuntos de descriptores que se definen para mejorar la organización lógica del tesoro. Estas agrupaciones se pueden hacer bien por campos (temas), o por clases de términos o "facetas" (categorías genéricas que agrupan descriptores que se pueden definir, por ejemplo, como fenómenos, procesos, materiales, organizaciones, seres vivos, equipamientos, propiedades, o disciplinas).
-

### b) *Relaciones semánticas*

Son los elementos que conforman la estructura del tesoro y según Arana y Codina (2004) se pueden clasificar en las siguientes categorías:

- Relaciones de equivalencia, que implican la existencia de términos preferentes y no preferentes, y por lo tanto son las que denotan el control terminológico en los tesauros. Implican la designación de términos preferentes o descriptores y otras formas posibles de representar un concepto (sinónimos y cuasi-sinónimos) que son descartadas como puntos de acceso en la indización y la recuperación.
  - Relaciones jerárquicas, que son las establecidas entre conceptos que presentan distintos niveles de superordinación y subordinación, y donde cada término subordinado debe tratar la misma clase básica de conceptos que su término general, por ejemplo representar un objeto, una acción o una propiedad. Existen cuatro subtipos de relaciones jerárquicas: género-especie (clase o categoría / miembros o especies), partitivas (parte/todo), enumerativas (categoría general / caso individual), y polijerárquicas (cuando un concepto se inserta en más de una cadena jerárquica y por lo tanto tiene más de un término genérico).
  - Las relaciones asociativas, son las más difíciles de definir y establecer, y en realidad se definen cuando una relación no puede identificarse como de equivalencia o jerárquica. La relación asociativa engloba, en realidad, varios tipos de relaciones, entre las que cabe destacar las siguientes:
    - Similitud (cuando se ha conferido la categoría de descriptores a términos que pueden ser considerados prácticamente sinónimos)
    - Causa y efecto
    - Instrumentalidad
    - Concomitancia
    - Sucesión en el tiempo o el espacio
-

- Relación con elementos constitutivos
- Relación de propiedad
- Objeto de una acción, proceso o disciplina
- Localización
- Antinomia

### **3.3.1.2 ONTOLOGÍAS**

Las ontologías web nos permiten dar un doble salto cualitativo con respecto a otras herramientas de organización del conocimiento: por un lado su uso permite pasar de la búsqueda en motores de búsqueda tradicionales con palabras clave a la recuperación de conceptos con la ayuda de agentes software, y por otro posibilitan recuperar información dentro de un contexto determinado, es decir, permiten recuperar conocimiento.

Los objetivos que persiguen, según Noy y McGuinness (Noy, McGuinness 2000) se podrían resumir en los siguientes puntos:

- a) Permitir la comprensión común de la estructura de la información de un dominio tanto por personas como por agentes software.
- b) Permitir la reutilización del conocimiento perteneciente a un dominio.
- c) Hacer explícitos los presupuestos (axiomas) y alcance de un dominio.
- d) Separar el conocimiento del dominio relativo a sus componentes del conocimiento operacional (relativo a los procesos o tareas que se desarrollan en ese dominio).
- e) Analizar el conocimiento de un campo a través de los términos y relaciones que lo configuran formalmente.

Atendiendo a su estructura es posible distinguir los siguientes componentes que son la pieza clave para representar el conocimiento de algún dominio (Gruber 1995):

---

- **Conceptos:** son las ideas básicas que se intentan formalizar. Los conceptos pueden ser clases de objetos, métodos, planes, estrategias, procesos de razonamiento, etc.
- **Relaciones:** representan la interacción y enlace entre los conceptos del dominio. Suelen formar la taxonomía del dominio. Por ejemplo: subclase-de, parte-de, parte-exhaustiva-de, conectado-a, etc.
- **Funciones:** son un tipo concreto de relación donde se identifica un elemento mediante el cálculo de una función que considera varios elementos de la ontología. Por ejemplo, pueden aparecer funciones como categorizar-clase, asignar-fecha, etc.
- **Instancias:** se utilizan para representar objetos determinados de un concepto.
- **Axiomas:** son asertos que se declaran sobre relaciones que deben cumplir los elementos de la ontología. Por ejemplo: "Si A y B son de la clase C, entonces A no es subclase de B", etc.

También es posible distinguir distintos tipos de ontologías de acuerdo a diferentes criterios. Guarino (1998) distingue tres tipos diferentes de ontologías atendiendo a su nivel de granularidad o especialización:

- **Ontologías de alto nivel:** Describen conceptos muy generales como espacio, tiempo, materiales, objetos, acontecimientos, etc. que son independientes de dominio.
  - **Ontologías de dominio y de tarea:** Describen, respectivamente, el vocabulario relacionado con un dominio genérico o con una actividad determinada, mediante la especialización de los términos definidos en una ontología de alto nivel.
  - **Ontologías de aplicación:** Describen conceptos que dependen tanto de una ontología de dominio como de una ontología de tarea a través de la especialización de los conceptos definidos en estas.
-



Corcho y colaboradores (2003) por su parte distinguen entre ontologías ligeras o *lightweight*, es decir, aquellas que presentan un bajo nivel de complejidad estructural, de las ontologías pesadas o *heavyweight*, que además añaden axiomas.

#### **3.4.1.3 TESAUROS VS. ONTOLOGÍAS**

Como vemos, tesauros y ontologías son herramientas bastante diferentes en cuanto a elementos, estructura y capacidades. Las ontologías son claramente superiores a los tesauros por varias razones (Qin, Paling 2001):

1. Las ontologías presentan un nivel más alto de descripción del vocabulario y un desarrollo semántico más profundo de las relaciones que establecen los conceptos entre sí.
2. Las ontologías se basan en la lógica descriptiva y favorecen el razonamiento inferencial de los agentes software.
3. Las ontologías son reutilizables y pueden trabajar en sistemas heterogéneos ya que permiten la descripción formal de objetos, sus propiedades y las relaciones que establecen entre ellos.
4. La información que contienen constituye auténticas bases de conocimiento.

No obstante, las diferencias entre tesauros y ontologías son cada vez más sutiles debido al progresivo enriquecimiento semántico al que se están viendo sometidos los nuevos tesauros que se desarrollan específicamente para su uso en la Web (Gilchrist 2003). Las relaciones que definen son cada vez más complejas con respecto a los tesauros tradicionales en papel o a los primeros tesauros desarrollados para la Web (generalmente en formato HTML) gracias a la aparición de vocabularios de etiquetado con una mayor capacidad expresiva como RDF Schema o SKOS Core (RDF Schema lo veremos en mayor profundidad en una próxima sección).

---

De hecho, SKOS Core es un esquema RDF creado específicamente para el diseño de sistemas de organización de conocimiento de una manera procesable por máquina dentro del ámbito de la Web Semántica cuyo principal objetivo es definir un modelo para migrar esquemas de conocimiento a la Web (Pérez Agüera 2004).

Esto provoca que la frontera entre los nuevos tesauros desarrollados para la web y las ontologías *lightweight* sea cada vez más difícil de delimitar.

### **3.4.2 TECNOLOGÍAS DE WEB SEMÁNTICA**

Como hemos visto en el capítulo 2, el proyecto Web Semántica es un escenario idóneo para el desarrollo de las bibliotecas digitales, y las tecnologías sobre las que se fundamenta elementos fundamentales que no solo permiten definir un vocabulario común interpretable por humanos y máquinas que favorece el intercambio y la interoperabilidad de la información, sino que ponen a nuestra disposición los medios necesarios para hacer representaciones de parcelas de la realidad para operar con ellas favoreciendo de esta forma el acceso a la información de una manera contextualizada. Esto convierte a las tecnologías de Web Semántica en una valiosa herramienta para el desarrollo de servicios de valor añadido.

En este apartado vamos a analizar algunos de los principales vocabularios de Web Semántica sobre las que se basa el proyecto Web Semántica y otros vocabularios que se han utilizado para desarrollar el modelo que presentamos en esta memoria. En concreto nos centraremos en el metalenguaje XML (base sintáctica del modelo de Web Semántica), RDF y RDF Schema (vocabularios que definen la infraestructura semántica de la Web Semántica), OWL (un lenguaje específico para definir ontologías), y RSS 1.0 y FOAF (aplicaciones específicas de RDF).

---

### 3.4.2.1 XML (EXTENSIBLE MARKUP LANGUAGE)

Uno de los elementos clave que forman la base sintáctica sobre la que se sustenta el modelo de Web Semántica es XML (*eXtensible Markup Language*), un metalenguaje que define una serie de normas básicas que permiten diseñar cualquier otro vocabulario de etiquetado para la Web.

Su origen se remonta a finales de los años noventa, en concreto a 1998, y surge como una exigencia desde determinados ámbitos (como el del comercio electrónico) de disponer de un lenguaje que ofreciera mayores capacidades que HTML para representar información. HTML es un lenguaje simple, fácil de usar, que permite definir la estructura espacial del documento a la hora de ser presentado en pantalla (estilo de las fuentes, párrafos, integración de elementos multimedia e interactivos, etc). El lenguaje dispone de un conjunto de etiquetas predefinidas que son fácilmente interpretables por las personas de manera que leyendo un trozo de código podemos, no solo conocer su contenido, sino que también es posible determinar que un bloque de texto específico será mostrado como un párrafo, o que determinada palabra aparecerá en negrita.

Sin embargo en este escenario la información no dispone de una estructura lógica explícita y las máquinas juegan un papel pasivo realizando prácticamente la misma tarea que un proyector, sin capacidad para comprender el significado de la información codificada en un documento HTML.

Un lenguaje con estas características y capacidades de hecho ya existía desde la década de los 80. El lenguaje estándar de etiquetado generalizado SGML (*Standard Generalized Markup Language*) (norma ISO 8879:1986) es un metalenguaje con el que cualquier individuo o institución puede diseñar su propio vocabulario sobre un dominio de interés para resolver un problema de representación de información determinado, y definir la estructura lógica de los documentos para que puedan ser interpretados no solo por humanos, sino también por máquinas (Moya, Peis 2000).

---

Sin embargo, SGML es un lenguaje demasiado general y complejo como para aplicarlo con éxito en un medio tan dinámico como la Web, donde la inmediatez, la transparencia y la facilidad de uso son los factores que determinan el éxito o fracaso de aplicaciones y lenguajes que pretenden ser ampliamente utilizados.

En 1998 aparece el lenguaje de etiquetado extensible XML (*eXtended Markup Language*) (Bray et al. 2006A]) como una simplificación de SGML que eliminaba las principales barreras que dificultaban la implantación exitosa de este en la Web.

El W3C definió una serie de prerequisites que debía de cumplir el nuevo lenguaje:

- Tenía que ser compatible con SGML.
  - Debía de ser un lenguaje directamente utilizable en Internet (es decir, ser legible en los navegadores web).
  - Debería servir de soporte para una amplia variedad de aplicaciones de transferencia de datos.
  - Era necesario que se los documentos pudieran ser procesados utilizando aplicaciones sencillas capaces de acceder a la estructura y el contenido de los documentos (como los *parsers* o analizadores sintácticos).
  - El número de características opcionales debería de ser mínimo, y a ser posible cero.
  - Los documentos deberían poder ser legibles por humanos y tener una estructura razonablemente clara.
  - La especificación del lenguaje debería de ser simple y concisa pero sin dejar de estar perfectamente formalizada.
  - Los documentos deberían de ser fáciles de crear por cualquier persona con herramientas simples (como un editor de texto).
-

Así, XML se define como un lenguaje de bajo nivel para intercambio de información estructurada entre distintas plataformas. Por lo tanto, el objetivo primordial de XML no es sustituir a HTML ya que ambos lenguajes están concebidos para desempeñar tareas diferentes: XML sirve exclusivamente para describir contenidos independientemente de la estructura del documento, mientras que HTML se utiliza para maquetar y presentar esos contenidos en pantalla.

Lo que sí es previsible es que a corto plazo se imponga el uso de vocabularios basados en XML (como es el caso de XHTML, una versión de HTML con sintaxis XML) ya que en los últimos años este estándar se ha extendido rápidamente en diferentes dominios como la empresa y la industria, o el ámbito científico y tecnológico, sirviendo de base sintáctica para el desarrollo de múltiples vocabularios especializados que sirven para resolver problemas específicos de esas áreas.

Gran parte del éxito de su implantación se fundamenta en las múltiples ventajas que supone su uso frente a otros vocabularios (Geroimenko, Chen 2003):

1. Es un estándar abierto definido por el W3C, independiente de plataforma y adoptado por la mayoría de empresas y distribuidores de software.
  2. XML utiliza UNICODE, un sistema de codificación multilingüe y universal de caracteres que facilita el intercambio y visualización de documentos independientemente del idioma (o alfabeto) en el que esté escrito.
  3. Proporciona un mecanismo simple y estandarizado para representar información (frente a otros vocabularios propietarios más complejos), lo cual agiliza su transmisión a través de la red y permite compartirla más fácilmente.
  4. La separación entre la representación formal y el contenido de los documentos permite reutilizar un mismo contenido en diferentes
-

aplicaciones (como bases de datos, hojas de cálculo, editores de texto o servicios web) y plataformas (teléfonos móviles, PDA's, ordenadores, etc.).

5. Los documentos XML contienen metadatos (información autodescriptiva) estructurada de manera jerárquica que los convierte en ricas fuentes de información.

Además, su sintaxis es realmente simple ya que la especificación del vocabulario se limita a establecer una serie de estrictas, aunque sencillas, reglas sintácticas para definir conjuntos de etiquetas (ya que no existe ningún conjunto preestablecido).

XML por lo tanto es un lenguaje extensible que permite a cada persona diseñar sus propias etiquetas lingüísticas para resolver un problema concreto de representación de información. Esto implica que la estructura de los documentos o la semántica de las etiquetas de cada uno de ellos se debe definir utilizando respectivamente esquemas sintácticos (DTD's o esquemas XML) y semánticos (como las ontologías).

Sin embargo, no todo son ventajas ya que la propia flexibilidad y capacidad expresiva de XML presentan por otro lado dos inconvenientes.

Por un lado XML no garantiza ni la interoperabilidad sintáctica entre sistemas heterogéneos. Cabe la posibilidad de que para resolver un mismo problema existan múltiples vocabularios, todos diferentes. Supongamos, por ejemplo, que tres individuos quieren diseñar un conjunto de etiquetas para describir personas y definen una etiqueta representar el documento nacional de identidad. Una de ellas crea la etiqueta `<dni>`, otro de ellos la etiqueta `<DNI>` y el tercero de ellos opta por definir la etiqueta `<documento_nacional_de_identidad>`. Estas tres etiquetas son fácilmente interpretables para los humanos: su contenido es una cadena alfanumérica de 8 cifras y una letra que identifica unívocamente a una persona nacida en España. Sin embargo, para un analizador sintáctico XML las tres etiquetas son completamente distintas ya que es incapaz de discernir cuál es

---

el contenido de estas. Para paliar este problema existen diferentes soluciones como el desarrollo de aplicaciones que permitan traducir unos vocabularios en otros, o la adopción de vocabularios a modo de convención común de comunicación que hayan sido diseñados por entidades de prestigio o autorizadas.

Por otro lado, XML tampoco es capaz de garantizar la interoperabilidad semántica ya que no dispone de mecanismos para determinar el significado de los elementos que describe. Para ello será necesario recurrir a vocabularios diferentes, como RDF (Beckett 2004) que establecen un modelos de datos capaz de representar la semántica de las etiquetas.

#### **3.4.2.1 FAMILIA DE ESPECIFICACIONES XML**

Existen docenas de vocabularios basados en esta sintaxis que permiten explotar las posibilidades de XML en aplicaciones específicas para ámbitos de conocimiento determinados. Estos vocabularios, hasta su aceptación definitiva como estándares de facto, sufren un proceso de maduración dentro del W3C. La primera propuesta del vocabulario se presenta en forma de informe técnico. Si este es aceptado pasa a borrador de trabajo donde comienza el desarrollo de la primera versión del vocabulario. El siguiente estadio es el de recomendación candidata y consiste en abrir un periodo de tiempo en el que se admiten sugerencias por parte de expertos para corregir errores, definir elementos y redefinir otros. De esta forma se obtiene una recomendación candidata que requiere una aprobación consensuada para convertirse definitivamente en una recomendación.

Las diferentes especificaciones se pueden dividir en tres grandes áreas (Geroimenko, Chen 2003):

- a) *Especificaciones para diseñar y crear lenguajes basados en XML (y sus correspondientes documentos)*
-

Dentro de esta categoría vamos a considerar al propio XML ya que, como hemos comentado anteriormente, se trata de un metalenguaje que permite diseñar cualquier conjunto de etiquetas, y por lo tanto va a ser la tecnología básica necesaria para desarrollar cualquier vocabulario de etiquetado orientado a su uso en la Web.

No obstante, va a necesitar de otras tecnologías auxiliares como, por ejemplo, los espacios de nombre o XML Namespaces (Bray et al. 2006B). Esta especificación, que no utiliza sintaxis XML, permite evitar la colisión entre nombres de elementos definiendo una referencia URI y un prefijo determinado para identificar cada conjunto de etiquetas. De esta forma podríamos conjugar en un mismo documento varias etiquetas homónimas sin riesgo a que el analizador sintáctico las considere como la misma etiqueta.

Otra tecnología básica para diseñar vocabularios con sintaxis XML es XML Schema (Fallside, Walmsley 2004). Permite definir esquemas sintácticos con los que definir la estructura y validar el contenido de diferentes tipos de documentos XML, definiendo qué elementos pueden ser incluidos y cuales no, la relación que existe entre ellos y el tipo de datos que pueden tomar como valor. De esta forma podemos hablar de documentos bien formados (cuando se ajustan exclusivamente a las reglas sintácticas de XML) y documentos válidos (cuando se ajustan a las reglas de un esquema). Son los sustitutos naturales de las DTD y en la actualidad se está ultimando el desarrollo de su versión 1.1 (Orchand2007).

XPath (Berglund et al. 2007) por su parte es otro vocabulario no XML con el que es posible especificar rutas para seleccionar nodos dentro de documentos XML de acuerdo a la estructura jerárquica de estos. A su vez, dentro de este vocabulario se incluye XPointer (DeRose, Maler, Daniel 2001), un vocabulario con el que es posible definir marcadores similares a las etiquetas A o anclas de HTML con los que, por ejemplo, se puede cargar en

---



un visualizador exclusivamente una serie de fragmentos seleccionados de un documento XML específico y que están identificados por estos marcadores.

También se incluyen en esta categoría otros vocabularios como XLink (DeRose, Maler, Orchard 2001), que permite definir vínculos complejos (bidireccionales, múltiples, etc.) entre recursos; XForms (Boyer et al. 2007) que es de utilidad para diseñar interfaces de usuario, y más específicamente, formularios web para realizar tareas de manipulación de datos simples; o XML Signature (Bartel et al. 2002), que permite crear y procesar firmas digitales, garantizando la integridad y autenticación de los recursos y de los servicios de firmas.

*b) Especificaciones para acceder, manipular, transformar y procesar documentos con sintaxis XML*

El segundo bloque de especificaciones lo forman aquellas con las que es posible manipular la información contenida en los documentos XML para someterla a procesos de edición, consulta o visualización, por citar solo algunos. El lenguaje que permite una modificación de los documentos a un nivel más profundo es DOM (*Document Object Model*)[168] ya que permite acceder directamente a la estructura jerárquica de los documentos XML para manipular sus nodos (pudiendo por ejemplo, añadir, editar o eliminar elementos).

Si lo que deseamos es transformar un documento XML en otro tipo de documento o bien modificar su formato de visualización disponemos de XSL (*eXtensible Stylesheet Language*) (Berglund 2006). Se compone de los sublenguajes XSL-T (*Transformation*) y XSL-FO (*Formatting Objects*) que permiten transformar documentos XML en documentos XML intermedios y estos a su vez en cualquier tipo de documento (doc, pdf, txt, etc.), y definir reglas de formato para su visualización en pantalla, respectivamente. Existe una alternativa a XSL-FO que consiste en el uso de hojas de estilo CSS (*Cascading Style Sheet Language*) (Bos et al. 2007). CSS es un vocabulario que

---

utiliza una sencilla sintaxis (no XML) que, aunque tiene una menor capacidad expresiva que XSL-FO, es apropiado para resolver problemas poco complejos de representación de información (por lo que es ampliamente utilizado dentro de la comunidad de desarrolladores web).

También está muy extendido dentro de esta comunidad el uso de XHTML (Pemberton et al. 2002), una reformulación de HTML 4.0 que combina la capacidad de descripción de XML con la versatilidad para representar información de HTML.

Otra forma de acceder al contenido de los documentos XML es mediante el uso de lenguajes de consulta, con los que es posible acceder a la información de los documentos XML de forma similar a como se consultan las bases de datos. La recomendación del W3C es XQuery (Boag et al. 2007).

Por último, en esta categoría también se incluyen otro tipo de lenguajes que permiten manipular los datos XML para su uso en diferentes aplicaciones que no trabajan en un ámbito textual. Así, por ejemplo, encontramos SVG (*Scalable Vector Graphics*) (Ferraiolo, Fujisawa, Jacson 2003), un estándar abierto que se utiliza para definir gráficos vectoriales en dos dimensiones, tanto estáticos como animados a partir de la información contenida en un documento XML, o VoiceXML (McGlashan et al. 2004), que se utiliza específicamente para desarrollar aplicaciones de reconocimiento de voz (como el software que se utiliza en las centralitas telefónicas o “*call centres*”).

c) *Especificaciones para añadir semántica a los recursos:*

En este grupo se aglutinan una serie de tecnologías cuyo objetivo es dar un paso más en la representación de información y añadir elementos que permitan contextualizarla, de manera que pasemos de representar información a representar conocimiento. A principios de esta década surge

---

XML Topic Maps (Pepper, Moore 2001), una especificación que permite crear redes semánticas a partir de las materias usadas para definir el contenido de varios documentos, y proporciona los medios para visualizarlas gráficamente en forma de mapas de conocimiento.

El otro gran grupo de especificaciones semánticas son las que se basan en el modelo de datos que define RDF (*Resource Description Framework*). Este vocabulario estructura la información en forma de tripletas sujeto-propiedad-objeto (identificadas unívocamente mediante referencias URI) de manera que esta se hace semánticamente accesible y puede ser procesada por agentes software para inferir nuevo conocimiento a partir de ella.

Sin embargo, para declarar la semántica de estos asertos es necesario disponer de mecanismos que permitan contextualizar los diferentes elementos descritos en los recursos RDF. Para cumplir esta misión surge RDF Schema (Brickley, Guha 2004), un vocabulario que extiende RDF y con el que se pueden definir una serie de relaciones semánticas básicas (como la pertenencia a clases, o el rango y dominio de propiedades).

En esta categoría también se incluye OWL (*Web Ontology Language*) (McGuinness, Harmelen 2004), un lenguaje que extiende las capacidades expresivas de RDF Schema para definir ontologías web, entendidas como la conceptualización de un ámbito de conocimiento a partir de una serie de conceptos relevantes de ese dominio y las relaciones semánticas complejas que establecen entre ellos.

#### **3.4.2.1.2 SINTAXIS BÁSICA**

Como acabamos de comentar, una de las principales bazas con las que ha contado XML para su rápida implantación es la sencillez de su sintaxis. Esta se podría resumir en una serie de reglas básicas, simples pero estrictas (para una información más completa, consultar la especificación -(Bray et al. 2006A))-:

---

- Todo documento XML se abre con una orden de procesamiento que incluye una serie de atributos que facilitan a los *parsers* la tarea de interpretar correctamente el documento XML. Los principales son los siguientes:
    - *version*: Es un atributo obligatorio en el que hay que consignar la versión de XML utilizada en el documento (en la actualidad solo existe la versión 1.0).
    - *encoding*: Aquí se indica el formato de codificación utilizada para diseñar el documento. Por defecto es UTF-8, aunque es posible utilizar otras, como UTF-16, US-ASCII, o ISO-8859-1.
    - *standalone*: Es un atributo opcional que se usa exclusivamente para definir si un documento tiene o no una DTD asociada. Sus valores válidos son “yes” (cuando el documento se considera *aislado* y no tiene asociada una DTD) y “no” (cuando sí que existe una DTD asociada). Este atributo no es aplicable cuando utilizamos XML Schemas.
  - Es preciso definir un único elemento raíz que contiene a todos los elementos necesarios para definir el recurso.
  - Los elementos que se utilizan en XML para describir recursos se organizan anidándose unos dentro de otros formando una estructura arbórea jerárquica.
  - Los elementos están definidos por etiquetas de inicio y fin de elemento delimitadas por los símbolos “<” y “>”. Ambas deben estar identificadas exactamente por la misma cadena de caracteres (XML es sensible a mayúsculas y minúsculas) excepto porque en la etiqueta de cierre del elemento el nombre está precedido por una barra inclinada. Entre ambas etiquetas se define el contenido, que puede ser una cadena de caracteres
-

(con ciertas restricciones sintácticas que comentamos más abajo) u otros elementos anidados.

- Existen determinados elementos que no tienen contenido en sí. La información que incluyen les viene dada por atributos. Se caracterizan porque el elemento se define mediante una única etiqueta que abre y cierra el elemento (se añade una barra de cierre justo antes del signo ">").
  - Los atributos permiten especificar el contenido de un elemento. Aparecen justo detrás del nombre del mismo, y es necesario definir su nombre, el signo "=", y a continuación, entre comillas dobles o simples, su valor (que debe ser una cadena de caracteres). Un elemento puede tener tantos atributos no repetidos como sea necesario, y se especifican únicamente dentro de las etiquetas de inicio de elemento y en los elementos vacíos.
  - Existe una serie de caracteres reservados para el procesamiento de los documentos XML que no pueden ser utilizados para definir el nombre de elementos y atributos o el contenido de los mismos. En concreto, no se pueden utilizar los caracteres " , ' & , < , > . En su lugar se usarán las entidades *&quot;*, *&apos;*, *&amper;*, *&lt;* y *&gt;* respectivamente.
  - Tanto el nombre de los elementos como el de los atributos siguen la misma sintaxis: pueden estar formados por cualquier cadena de caracteres alfanumérica, excepto por espacios en blanco, caracteres ortográficos especiales (tildes, ñ, ç, /, etc.) o los caracteres reservados de XML.
  - El contenido de elementos y atributos puede contener cualquier cadena de caracteres excepto los caracteres reservados de XML.
-

Con estas normas básicas ya podríamos definir documentos XML bien formados, que son aquellos que cumplen las reglas sintácticas del vocabulario pero sin estar necesariamente sujetos a ninguna otra regla sobre la estructura, ocurrencia, orden o valor de elementos y atributos que contienen.

Si por el contrario, además de estar bien formado, en el documento se definen los elementos que lo componen conforme a una serie de restricciones estructurales definidas en un esquema sintáctico, como las DTD (*Document Type Definitions*) o los XML Schemas, (y que pueden ser procesadas por un *parser* o analizador sintáctico) entonces podemos decir que el documento XML es un documento válido.

#### **3.4.2.1.3 TECNOLOGÍAS AUXILIARES**

No obstante, y a pesar de que XML se conforma como la base sintáctica de la Web Semántica, necesita a su vez apoyarse en otras dos tecnologías web auxiliares: Unicode y las referencias URI (*Uniform Resource Identifier*).

Unicode es un vocabulario que codifica los caracteres de los idiomas más comunes utilizando dígitos enteros positivos, y define diferentes codificaciones binarias de estos caracteres (siendo las más comunes UTF8, UTF16, UTF32, e ISO-8859-1 que codifica los caracteres latinos).

Por su parte, las referencias URI permiten asignar un identificador único a los recursos web de forma que quedan caracterizados de manera unívoca.

Los elementos básicos que conforman una URI son los siguientes (Berners-Lee, Fielding, Masinter 2005):

---

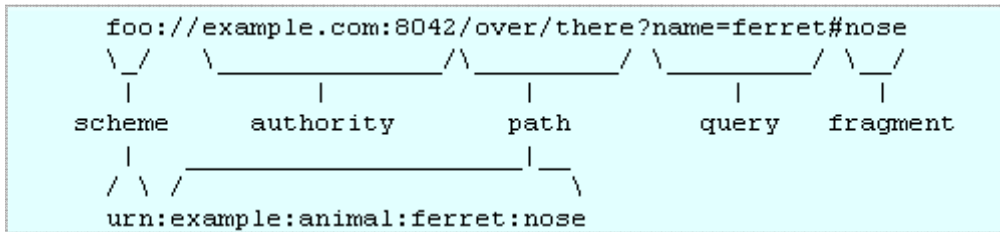


Fig.15 Esquema de componentes básicos de una URI

- El elemento *esquema* define un prefijo que identifica la especificación utilizada para asignar los identificadores (como las *urn:*) o incluso el protocolo de acceso a los recursos (*http:*, *ftp:*, *mailto:*, etc.).
- El elemento *autoridad* es opcional, y se corresponde con un elemento jerárquico, denominado autoridad de nombres, que identifica un dominio web (como por ejemplo, la dirección de un servidor web).
- El elemento *ruta* identifica un camino jerárquico al recurso dentro del ámbito de la autoridad de nombres.
- El elemento *consulta* va siempre precedido por un elemento *ruta* y permite identificar un recurso utilizando información no jerárquica. Este elemento se introduce mediante el carácter “?” y generalmente viene definido por un par “clave=valor”.
- El elemento *fragmento* permite identificar una parte específica del recurso y viene precedido por el carácter “#”.

No obstante la forma más común que suelen adoptar las URI's es la de las URL's (aunque esta no tiene por qué apuntar necesariamente a un recurso almacenado físicamente en esa dirección).

Las URI's utilizadas como identificadores únicos no solo permiten identificar recursos sino que también son un elemento esencial para la definición de espacios de nombre XML (Bray et al. 2006B). Los espacios de nombre

representan un mecanismo para identificar vocabularios de etiquetado (es decir, conjuntos de etiquetas). De esta manera en un único documento XML sería posible utilizar simultáneamente diferentes vocabularios sin que haya colisión entre los elementos y atributos que ambos definen (siempre que estén todos convenientemente identificados mediante su correspondiente espacio de nombre y prefijo).

Para aclarar este concepto pongamos un ejemplo concreto. Supongamos que estamos definiendo un documento en el que se describe el mobiliario que se va a utilizar para equipar una sucursal bancaria. A parte del vocabulario que definimos nosotros para realizar la descripción, vamos a necesitar otros dos conjuntos de etiquetas: uno que defina elementos descriptivos para entidades bancarias y otro sobre mobiliario.

Identificando adecuadamente cada conjunto con su respectivo espacio de nombre en el ámbito de un elemento determinado (por lo general en el elemento raíz) es posible utilizar etiquetas de cada uno de ellos sin que exista riesgo de colisión, aunque existan etiquetas homónimas (como es el caso de las etiquetas “*banco*” del ejemplo).

```
<?xml version="1.0"?>
  <eq:equipamiento xmlns:eq="http://equipment.net"
                  xmlns:bank="http://economy.org/1999/"
                  xmlns:furn="http://furniture.com/">
    <bank:banco>La Caixa</bank:banco>
    <eq:mobiliario>
      <furn:mesa_escritorio modelo="Alabama-X44" />
      <furn:banco modelo="Nogal 12-country" />
    </eq:mobiliario>
  </eq:equipamiento>
```

Fig.16 Ejemplo del uso de espacios de nombre en XML



### 3.4.2.2 RDF y RDF Schema

El vocabulario básico sobre el que se sustenta el modelo de Web Semántica es RDF (*Resource Description Framework*) y como su propio nombre indica, define un modelo de datos que permite describir en la Web recursos, objetos físicos, conceptos abstractos o cualquier otra cosa con identidad propia.

La descripción se realiza en términos de las propiedades que caracterizan a los recursos y los valores que estas toman, de forma que es posible enriquecer la definición de un recurso simplemente añadiéndole nuevas propiedades. Por esta razón es posible afirmar que RDF es un vocabulario centrado en propiedades (McBride 2004).

El modelo de datos RDF define la forma en que debe ser estructurada la información sobre un marco conceptual abstracto cuya unidad semántica mínima es la sentencia o declaración.

Cada recurso está pues formado por al menos una sentencia, y esta a su vez por una tripleta *sujeto-propiedad-objeto*, donde el sujeto indica a qué (o quién) se refiere la sentencia, la propiedad indica una característica determinada del sujeto, y el objeto indica el valor que toma la propiedad.

Los recursos RDF se pueden representar en forma de grafos orientados con nodos y vértices etiquetados que son interpretados de acuerdo a las siguientes convenciones:

- Los nodos elípticos representan los recursos y están identificados por una URI que los identifica de manera unívoca, sea cual sea su naturaleza e independientemente de su ubicación física.
  - Los nodos elípticos vacíos representan elementos que no están identificados por una URI. La manera de interpretar la presencia de un nodo anónimo es que existe al menos un recurso que presenta las propiedades y valores definidos para el nodo.
-

- Los vértices representan propiedades del recurso. Son características o atributos específicos (identificados por su propia referencia URI) que permiten describir recursos y que pueden tomar como valor un literal u otro recurso.
- Los nodos rectangulares representan valores literales con o sin tipo de datos asociado.

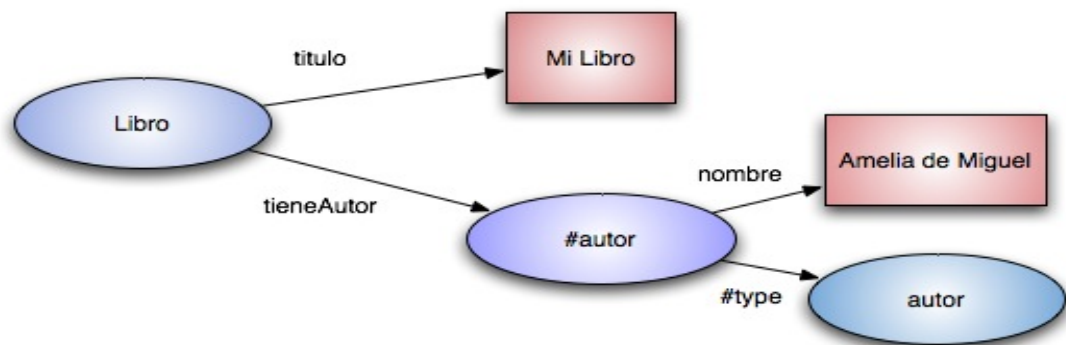


Fig.17 Grafo RDF

Con objeto de facilitar el procesamiento y dentro de lo posible la comprensibilidad del grafo se suele recurrir a diferentes sintaxis para definir los recursos. La más extendida es la sintaxis XML, aunque existen otras sintaxis alternativas que inciden en clarificar y facilitar la interpretación de los asertos RDF a las personas, como Notation 3 (también denominada N3) (Berners-Lee, 2006), o N-triples (Grant, Beckett 2004), una sintaxis específica para estudios de casos donde las tripletas aparecen listadas secuencialmente, de forma que es más simple seguir y rastrear razonamientos e inferencias.

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
        xmlns:ex="http://ejemplo.org/">
  <rdf:Description rdf:about="http://example.org/Libro">
    <ex:titulo>Mi libro</ex:titulo>
    <ex:tieneAutor>
      <ex:autor rdf:ID="autor01">
        <ex:nombre>Amelia de Miguel</ex:nombre>
      </ex:autor>
    </ex:tieneAutor>
  </rdf:Description>
</rdf:RDF>
```

Fig.18 Representación del grafo de la *fig.17* con sintaxis XML

No obstante, y a pesar de que RDF define una infraestructura para la representación de conocimiento, no tiene capacidad para expresar la semántica del vocabulario (es decir, el significado de los sujetos, predicados y objetos definidos, y las relaciones que establecen entre ellos). En otras palabras, RDF no dispone de ningún mecanismo de razonamiento sobre los asertos que define.

Por esta razón, para definir el significado de una sentencia RDF es necesario definir un esquema semántico que permita contextualizar las sentencias de manera que el autor del recurso y el receptor del mismo entiendan el mismo significado. Los esquemas semánticos son como diccionarios que definen los elementos que se pueden utilizar en una sentencia y a los que les asigna un significado específico.

El esquema semántico de RDF es RDF Schema (Brickley, Guha 2004), un vocabulario que se define como extensión semántica de RDF, que aprovecha su modelo de datos y gran parte de los elementos de modelado pero que introduce

---

algunas herramientas ontológicas simples que permiten, por ejemplo, definir clases de elementos y propiedades (así como jerarquías dentro de ambas).

Por esta razón algunos autores se refieren a ambos lenguajes como un único vocabulario que denominan RDF(S). Adoptando este criterio, vamos a hacer un breve repaso sobre la sintaxis de ambos vocabularios de una manera integrada (para mayor información, consultar las especificaciones correspondientes - (Beckett 2004) (Brickley, Guha 2004)-).

#### **3.4.2.2.1 SINTAXIS BÁSICA**

Todo documento RDF básicamente se compone de un conjunto de recursos, un subconjunto de recursos denominados “propiedades” y un conjunto de literales. Además también se definen una serie de sentencias formadas por tripletas sujeto-predicado-objeto (recurso, propiedad, recurso/literal).

Los recursos son instancias de la clase *rdfs:Resource* y para describirlos se utiliza el elemento *rdf:Description*, en cuyo ámbito se declaran las diferentes propiedades que lo caracterizan con sus valores correspondientes. Cada elemento está identificado por su correspondiente referencia URI que se define mediante el atributo *rdf:about*.

En lugar del atributo *rdf:about* también es posible utilizar el atributo *rdf:ID* cuando lo que deseemos sea necesario abreviar URI's complicadas o de uso frecuente. En este caso la URI del recurso es construida por el *parser* tomando como base la referencia URI definida en un atributo *xml:base* y añadiendo el identificador único consignado en *rdf:ID*.

También es posible identificar nodos anónimos mediante identificadores únicos aunque no necesariamente mediante una referencia URI, como por ejemplo ocurre cuando estamos describiendo un recurso que no está en la Web (un objeto físico). En este caso el elemento *rdf:Description* va acompañado del

---

atributo *rdf:nodeID* cuyo valor puede ser utilizado en para referenciar el recurso en cualquier parte del documento RDF.

#### A) CLASES Y SUBCLASES

RDF Schema amplía la sintaxis básica de RDF añadiendo nuevos elementos como las clases. Estas se definen como mecanismos de abstracción que permiten agrupar diferentes recursos que presentan características similares (Antinou, van Harmelen 2004B), ya sean entidades, personas o cosas. A partir de esta definición es posible introducir el concepto de extensión de clase, entendida como el conjunto de individuos o instancias que forman parte de una clase determinada.

No obstante, las instancias en sí no son las que dan el significado a la clase ya que dos clases distintas pueden tener la misma extensión de clase (por ejemplo, la clase “*Años de la década de los noventa*” y la clase “*Años de mandato*” de un cargo determinado pueden llegar a tener exactamente la misma extensión de clase aunque la semántica de cada una de ellas es claramente distinta). La semántica viene dada por el contexto y la relación de cada clase con el resto de clases de un dominio.

En RDF, por la propiedad de autorreferencia las clases (*rdfs:Class*) tienen la particularidad de que pueden ser consideradas a un tiempo como recursos y como miembros de sí mismas como clase. No obstante, para evitar violar los principios de la teoría clásica de conjuntos de Zermelo y Fraenkel (Zermelo 1930) en ámbitos determinados se suele distinguir de manera inequívoca cuando la clase actúa de una u otra manera.

Para declarar recursos como miembros de una clase es necesario utilizar dentro del ámbito de un elemento *rdf:Description* la propiedad *rdf:type* que define el “tipo de recurso” que estamos describiendo mediante la adscripción del mismo a una clase determinada a través de una referencia URI.

---

Existe otra manera de describir elementos de clase sin necesidad de utilizar *rdf:Description* ni *rdf:type*. Consiste en declarar elementos “tipificados” en los que se define como nombre de elemento el del recurso que forma parte de la clase y al que se le asigna un identificador único utilizando el atributo *rdf:ID*.

```
<rdf:Description rdf:about="http://example.org/Libro">
    <rdf:type rdf:resource="http://www.URI.net/TipoElemento"/>
-----
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xml:base="http://example.org/">
    <TipoElemento rdf:ID="Libro">
```

Fig.19 Declaración de elementos de clase

En RDF(S) también se pueden definir jerarquías entre clases utilizando la propiedad *rdfs:subclassOf*, con la que es posible declarar que para dos clases determinadas las instancias de una son a su vez instancias de la otra.

## B) REIFICACIÓN

Es una propiedad de RDF específica para representar información en la Web que permite rastrear el origen de un aserto RDF. De esta manera es posible comprobar y garantizar la corrección, veracidad, integridad y confiabilidad de la información.

Las sentencias RDF *reificadas* son recursos que representan la ocurrencia de una sentencia RDF determinada. Estos recursos (que pertenecen a la clase *rdfs:Statement*) presentan tres propiedades básicas: *rdf:subject*, *rdf:predicate* y *rdf:object* que representan respectivamente a sujetos, predicados y objetos de la sentencia para crear una nueva sentencia completamente distinta a la que describe, y a la que es posible añadir otras propiedades que permitan especificar

la semántica del recurso (como la autoría de la misma, el momento en que fue creada, el lugar, etc.).

No obstante, y dada su gran expresividad, la reificación supone una potente herramienta para describir información, pero dificulta el desarrollo de sistemas de razonamiento eficientes.

### C) CONTENEDORES

Los contenedores son elementos pertenecientes a la clase *rdfs:Container* que permiten agrupar un número ilimitado de recursos como valores de una propiedad de pertenencia específica.

Existen tres tipos de contenedores. Por un lado, encontramos las *rdf:Bag* que definen agrupaciones de recursos en los que el orden en que aparecen estos no es relevante. Los *rdf:Seq* por su parte agrupan recursos donde el orden sí que es significativo, mientras que los *rdf:Alt* permiten definir agrupaciones de recursos de entre los cuales es posible seleccionar uno de forma excluyente.

No obstante, las características propias de cada una de ellas son una convención que no está definida explícitamente en la semántica de RDF, por lo que la diferenciación entre los tres tipos de contenedores viene impuesta exclusivamente por el uso que de ellos se haga.

Los contenedores por lo general están identificados mediante un *rdf:ID* único y pueden contener listados más o menos estructurados que se definen mediante propiedades *rdf:li*, (de manera similar a como se hace en HTML) cuando exclusivamente queremos especificar una secuencia de elementos de lista, o utilizando propiedades *rdf:\_N* identificadas por números correlativos enteros positivos mayores que cero (“N”) que definen el orden que ocupa cada uno de los elementos dentro del contenedor.

---

#### D) LISTADOS

Existe una clase genérica *rdf:List* que permite el diseño de listados (que se utiliza sobre todo como base para otros lenguajes que extienden la semántica de RDF(S)). Dentro de esta clase se definen las propiedades *rdf:first* para identificar al primer elemento del listado, *rdf:rest* para identificar cualquier elemento que no es el primero del listado, y *rdf:nil* para especificar listas vacías.

#### E) PROPIEDADES Y SUBPROPIEDADES

Las propiedades se definen en RDF como instancias de la clase *rdf:Property* y se utilizan para definir aspectos determinados que caracterizan a una o varias clases. Como instancias de una clase, se pueden autorreferenciar, y también es posible definir para ellas jerarquías con la propiedad *rdfs:subPropertyOf*.

Como sabemos, el valor de las propiedades puede ser un literal o un recurso. En el caso de los literales cabe la posibilidad de declarar su tipo de dato correspondiente mediante el constructor *rdf:datatype*. Admite el uso de los diferentes tipos de datos nativos de XML Schema y de nuevos tipos obtenidos como resultado de la extensión o restricción de estos. Todos ellos sin excepción se incluyen a su vez como miembros de la clase *rdfs:Datatype* que contiene todos los tipos de dato que se puedan definir en un documento RDF.

Para definir que el valor del objeto de un predicado es un recurso se puede recurrir al atributo *rdf:resource* que declara, dentro del ámbito de la propiedad, la referencia URI del recurso que toma como valor.

---



## F) PROCESAMIENTO DE PROPIEDADES

RDF dispone de diferentes mecanismos para procesar las propiedades definidas en un recurso de manera distinta a la habitual, dotándolas de nuevas características. Esto se consigue mediante el atributo *rdf:parseType*, el cual puede tomar tres valores diferentes: *Literal*, *Resource* y *Collection*.

RDF de manera nativa admite al uso de algunos elementos de XML, como por ejemplo el atributo *xml:lang* para especificar el idioma en el que está expresado el valor de una propiedad determinada, pero no permite la inclusión de etiquetas XML porque no siguen el modelo de datos de RDF. El único modo de incluirlas es procesándolas de manera que sean interpretadas por el analizador sintáctico como valor literal de una propiedad determinada. Esto se consigue definiendo en dicha propiedad el atributo *rdf:parseType* con valor *Literal*.

```
<rdf:Description rdf:about="http://examp.org/item01">
  <ex:prop rdf:parseType="Literal" xmlns:a="http://examp.org/a#">
    <a:elementoXML atributo="valor"/>
  </ex:prop>
</rdf:Description>
```

Fig.20 Ejemplo de uso del atributo *rdf:parseType="Literal"*

El valor *Resource* se utiliza para definir elementos anónimos (aquellos que no están identificados por una URI) sin que sea necesario definir el constructor *rdf:Description*. De esta manera es posible convertir una propiedad ordinaria en una propiedad-nodo que puede contener otras propiedades y atributos (excepto el atributo *rdf:ID* que implica la asociación de una referencia URI).

En este tipo de propiedades-nodo es habitual definir valores cualificados (por ejemplo cuando se cuantifica una dimensión usando una unidad determinada). Es estos casos para la propiedad se define un atributo *rdf:value*.

---

```
<rdf:Description rdf:about="http://ejemplo.com/#Adam_Wright">
  <tal:peso rdf:parseType="Resource">
    <rdf:value rdf:datatype="xsd:decimal">80.4</rdf:value>
    <tal:unidad rdf:resource="http://unidad.org/kilos"/>
  </tal:peso>
</rdf:Description>
```

Fig.21 Ejemplo de uso de *rdf:value*

El valor “Collection” de *rdf:parseType*, por su parte, permite definir conjuntos de recursos como valor de una propiedad. Las colecciones, a diferencia de otras agrupaciones de elementos como los contenedores, definen grupos cerrados que admiten un número finito de miembros.

#### G) DOMINIO Y RANGO

Una de las herramientas más útiles que incorpora RDF Schema al vocabulario es la posibilidad de definir relaciones de pertenencia de sujetos y objetos de una propiedad específica a una o varias clases. Para ello cuenta con dos propiedades: *rdfs:domain* que permite declarar que todos los sujetos de una propiedad pertenecen a una clase determinada (dominio), y la propiedad *rdfs:range* con la que se es posible definir la pertenencia de los objetos de una propiedad determinada a una clase específica (rango).

#### H) PROPIEDADES DE DEFINICIÓN

En RDF(S) existe un tipo especial de propiedades que permiten especificar y clarificar la semántica de un recurso determinado añadiendo información suplementaria. Son las denominadas propiedades de definición, de las que se distinguen dos clases: las orientadas al usuario y las orientadas a las máquinas.

Dentro de las orientadas al usuario encontramos *rdfs:label* y *rdfs:comment*. En concreto *rdfs:label* permite definir una denominación legible por el usuario (una etiqueta lingüística) que le facilite la identificación de recursos caracterizados a través de una referencia URI. Por su parte *rdfs:comment* permite aportar información adicional (añadir una nota o comentario) sobre el recurso en forma de bloque de texto no procesable por máquina.

Dentro de las propiedades de definición orientadas a las máquinas encontramos *rdfs:seeAlso*, que permite hacer referencia a un recurso que aporta nueva información sobre el recurso que está siendo descrito, y *rdfs:isDefinedBy*, una sub-propiedad de *rdfs:seeAlso* que por lo general se utiliza para declarar el RDF Schema que define la semántica de un recurso RDF.

#### I) LIMITACIONES DE RDF(S)

RDF(S) no soporta conceptos ontológicos básicos como equivalencia, disyunción, negación, relaciones inversas, y limitaciones de cardinalidad que sí están presentes en otros lenguajes. De esta forma se asegura la interoperabilidad entre lenguajes que se definan como extensión semántica de RDF Schema.

RDF(S) es un lenguaje de expresividad deliberadamente limitada (RDF define predicados y RDF Schema jerarquías de clases y propiedades, y propiedades de rango y dominio). Por ejemplo, con RDF(S) no es posible definir localmente rangos de propiedades, ni relaciones de disyunción entre clases (solo jerarquías de clases). No dispone de herramientas para realizar combinaciones booleanas de clases, ni para definir restricciones de cardinalidad (número de valores

---

diferentes que puede tomar una propiedad determinada). Tampoco es posible definir propiedades con características especiales como las propiedades transitivas (por ejemplo, *es\_madre\_de*), propiedades de unicidad (por ejemplo, *tiene\_DNI*) o propiedades inversas (por ejemplo, *es\_hijo\_de* definido como propiedad inversa de *es\_madre\_de*).

Para suplir esta falta de expresividad y cubrir los requerimientos para diseñar ontologías se han desarrollado diferentes lenguajes ontológicos que disponen de mayores capacidades que RDF(S) para definir ontologías.

---

### 3.4.2.3 FOAF (FRIEND OF A FRIEND)

El vocabulario *Friend Of A Friend* (FOAF) (Bricley, Miller 2005) (cuya traducción al español sería “*amigo de un amigo*”) surge como un esfuerzo de colaboración entre diferentes desarrolladores de software involucrados en el proyecto Web Semántica para crear un vocabulario que facilitara la integración y el intercambio de información personal entre diferentes aplicaciones. Para ello, el vocabulario adopta la sintaxis de XML y el modelo de datos de RDF para definir una ontología OWL en la que se incluye la definición de los diferentes elementos que lo componen y la relación que existe entre ellos, facilitando de esta forma a las aplicaciones operar con ellos y con las descripciones de recursos.

La filosofía que subyace al vocabulario (y que ya se intuye en su propio nombre) es considerar a las personas como los elementos básicos que conforman la propia Web, que establecen redes sociales entre ellos y que están relacionados con los recursos que están descritos en la Web, bien como autores de estos recursos (relación entre un fotógrafo y las fotografías que toma), bien como objeto de ellos (las personas que aparecen en las fotografías). De esta manera, a la hora de describir a una persona es posible hacer referencia además a los recursos con los que el individuo tiene algún tipo de nexo o relación.

#### 3.3.2.3.1 SINTAXIS BÁSICA

FOAF dispone de un núcleo básico estable de elementos y propiedades que permiten describir personas y las tareas que desempeñan sin imponer restricciones sobre el enriquecimiento de la descripción utilizando otros vocabularios RDF mezclados de forma eficiente y consistente, como SKOS Core (Miles, Brickley 2008) para definir materias, o microformatos como GRDDL (Connolly 2007) o RDFa (Adida, Birbeck 2007).

El núcleo de FOAF está compuesto por cinco clases principales de elementos entre las que se distribuyen un total de 12 clases y 54 propiedades. En la primera

---

de las categorías se agrupan clases y propiedades básicas que caracterizan a una persona en la Web, como su nombre y apellidos, su sitio web personal o su correo electrónico. La segunda de estas categorías contiene elementos para describir información personal adicional como el lugar de trabajo, la dirección del blog personal del individuo o las materias que son de su interés.

Otra de las categorías recoge información relativa a las cuentas en línea que puede mantener el individuo (abiertas en tiendas en línea, sitios de subastas, de apuestas, etc.), o los identificadores de cuentas de mensajería instantánea (*chat*) más comunes, como ICQ o MSN Messenger. La categoría “*Documentos*” contiene básicamente elementos que permiten asociar de una manera rudimentaria al individuo descrito con aquellos recursos de los que es autor o en los que aparece (como, por ejemplo, una fotografía o un artículo periodístico). La última de las categorías define metadatos sobre los grupos y proyectos de cualquier naturaleza en los que se encuentre involucrado el individuo.

Dada la gran cantidad de elementos que conforman el vocabulario, en este apartado nos vamos a limitar a hacer una breve revisión de los que se utilizan con más frecuencia.

El vocabulario define 4 clases principales:

- *Agent* es la clase que engloba aquellas entidades que tienen la capacidad de realizar alguna acción. Incluye a su vez tres subclases que caracterizan a personas (*Person*), grupos sociales (*Group*), y organizaciones (*Organization*). Los grupos suelen estar formados por personas por lo que dentro de la clase *Group* habría que indicar este hecho haciendo referencia a la clase *Person* mediante reglas OWL dentro de la propiedad *membershipClass* (para definir exactamente qué clase de personas forman el grupo), o bien usar la propiedad *member* para listar una a una las personas que lo componen. Por su parte la clase *Organization* es bastante similar a *Groups* aunque quizás su definición sea más consistente que la de grupo, ya que sus miembros (sociedades,
-

organizaciones, instituciones, compañías, etc.) definen un marco social más concreto que una agrupación genérica de personas.

- La clase *Document* permite describir cualquier recurso presente en la Web que se pueda considerar como tal, y que de hecho mantiene algún tipo de vínculo con un agente de la Web (es decir, con una persona, un grupo o una organización). Presenta a su vez dos subclases: *Image*, que permite describir imágenes, y *PersonalProfileDocument* que sirve para definir propiedades sobre la persona que es autora (*maker*) de un documento específico del que, a su vez, se puede considerar su materia principal (*primaryTopic*). En nuestro caso, esta subclase será de utilidad para caracterizar los perfiles personales que los propios usuarios crean sobre sí mismos.
- La clase *Project* permite especificar iniciativas de cualquier naturaleza (ya sean formales, informales, individuales o colectivas). Las instancias de esta clase suelen tener asociado un sitio web (representado por la propiedad *homepage*) donde debería aparecer la documentación relativa al proyecto.
- *OnlineAccount* es una clase que permite indicar cualquier servicio en línea proporcionado a un agente determinado por parte de un proveedor web específico (y que se identifica mediante la propiedad *accountServiceHomepage*). Existen diferentes tipos de cuentas que están representadas por sus correspondientes subclases como cuentas de mensajería instantánea (*OnlineChatAccount*), cuentas comerciales (*OnlineEcommerceAccount*) o cuentas en sitios web de juego o apuestas (*OnlineGamingAccount*).

Las propiedades que se definen en el vocabulario se agrupan en 5 grandes categorías (tal y como aparecen en la tabla 1):

---

ELEMENTOS FOAF BÁSICOS	CUENTAS EN LÍNEA Y MENSAJERÍA INSTANTÁNEA
Agent Person name nick title homepage mbox mbox_sha1sum phone img gender depiction (depicts) surname familyname givenName firstName birthday	OnlineAccount OnlineChatAccount OnlineEcommerceAccount OnlineGamingAccount holdsAccount accountServiceHomepage accountName icqChatID msnChatID aimChatID jabberID yahooChatID
INFORMACIÓN PERSONAL	Grupos y proyectos
weblog knows interest currentProject / pastProject plan based_near workplaceHomepage workInfoHomepage schoolHomepage topic_interest publications geekcode myersBriggs dnaChecksum	Project Organization Group member/ membershipClass fundedBy theme
	Documentos e imágenes
	Document Image PersonalProfileDocument topic (page) primaryTopic (isPrimaryTopicOf) tipjar sha1 made (maker) thumbnail logo openid

Tabla 1. Clases y propiedades de FOAF



- *Propiedades FOAF básicas:* Aquí aparecen agrupadas propiedades que permiten definir los datos básicos con los que se puede caracterizar a una persona como, por ejemplo, su nombre (*name, surname, firstname, givenname, family\_name, nick*), su fecha de nacimiento (*birthday*), su web personal o cuenta de correo electrónico (*homepage, m\_box*), o su imagen (*img, depiction*).
  - *Propiedades sobre información personal sensible:* Recoge propiedades con las que es posible definir datos más específicos sobre la persona, como sus intereses personales (*interest, topic\_interest*), sus proyectos (*currentProject, pastProject*), sus publicaciones (*publications*), su blog (*weblog*), información sobre su lugar de trabajo (*workplaceHomepage, workInfoHomepage*), localizaciones geográficas asociadas a la persona (*based\_near*), o incluso especificar a qué otras personas conoce (propiedad *knows*), lo cual permite definir redes sociales entre los usuarios de la Web.
  - *Propiedades sobre cuentas en línea y mensajería instantánea:* Este grupo contiene propiedades que permiten caracterizar e identificar las diferentes cuentas que puede tener una persona (*holdsAccount, accountName, accountServiceHomepage*), y los identificadores de diferentes servicios de mensajería instantánea (*yahooChatID, icqChatID, etc.*).
  - *Propiedades sobre grupos y proyectos:* Las propiedades de este grupo permiten identificar tanto la clase de individuos que forman parte de un grupo como los miembros del mismo (*membershipClass, member*), o las fuentes de financiación de los proyectos (*fundedBy*).
  - *Propiedades sobre documentos e imágenes:* Entre otras, en este grupo aparecen propiedades que permiten definir el autor de un documento o
-

imagen (*maker*), o las materias que describen su contenido (*topic*, *primaryTopic*).

En nuestro sistema utilizamos FOAF para definir la parte del perfil del usuario que describe sus datos personales de acuerdo a la siguiente estructura (de la que hemos obviado algunos elementos que no pertenecen a FOAF):

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:foaf="http://xmlns.com/foaf/0.9/"
  ...>
<foaf:PersonalProfileDocument rdf:about="">
  <foaf:maker rdf:resource="#persona"/>
  <foaf:primaryTopic rdf:resource="#persona"/>
</foaf:PersonalProfileDocument>
<foaf:Person rdf:ID="persona">
  <foaf:name>Diego Allione</foaf:name>
  <foaf:title>Sr.</foaf:title>
  <foaf:givenname>Diego</foaf:givenname>
  <foaf:family_name>Allione</foaf:family_name>
  <foaf:homepage rdf:resource="http://allione.org"/>
  <foaf:depiction rdf:resource="allione.jpg"/>
  ...
</foaf:Person>
</rdf:RDF>
```

Fig.22 Ejemplo de perfil de usuario en formato FOAF

Utilizamos básicamente la subclase *PersonalProfileDocument* y la clase *Person*. Dentro de *PersonalProfileDocument* definimos las propiedades *maker* y *primaryTopic*, ambas referenciando la descripción del usuario definida dentro de

la clase *Person*. Dentro de esta clase definimos varias propiedades para identificar al usuario como *name*, *title*, *givenname*, o *family\_name*, su página web personal (*homepage*), o una imagen que lo represente (*depiction*).

#### **3.4.2.4 RSS 1.0 (RDF SITE SUMMARY)**

El vocabulario RSS 1.0 (Begeed-Dove et al. 2001B) es, junto a otros lenguajes como RSS 2.0 o Atom, una de las tecnologías sobre las que se apoya el desarrollo de servicios de redifusión de contenidos. Es un vocabulario de marcado especialmente diseñado para gestionar listados de hiperenlaces que hace posible la publicación y difusión de contenidos web de una manera sencilla y flexible. Supone una opción para navegar por múltiples sitios web, consultando información actualizada, sin la necesidad de visitar todos y cada uno de esos sitios.

Sus principales objetivos son:

- definir un vocabulario sencillo fácil de usar por cualquier persona;
  - se puede utilizar para diversos fines, como presentar listados de datos o catalogar documentos, por citar solo algunos de ellos;
  - es posible extender el vocabulario mediante módulos para resolver problemas concretos sin necesidad de modificar el núcleo central del mismo;
  - el uso de metadatos permite crear descripciones enriquecidas de los recursos.
  - facilita la publicación y difusión de contenidos en la red, así como su posterior agregación para su presentación a los usuarios.
-

#### 3.4.2.4.1 SINTAXIS BÁSICA

El núcleo del vocabulario permite describir los 4 elementos informativos que son componentes básicos de todo canal RSS:

- Canal *<channel>*: En este elemento obligatorio es posible describir el canal en sí, incluyendo su título, el enlace a la página web a la cual está asociada, un pequeño resumen sobre su contenido, y un listado de los ítems que contiene (a modo de tabla de contenidos). Adicionalmente, se pueden referenciar el resto de elementos optativos que pueden formar parte del canal (imagen, y entrada de texto).
- Imagen del canal *<image>*: Opcionalmente, es posible asociar una imagen al canal que será descrita en función de su título, URL donde se encuentra físicamente el fichero de imagen y URL a la que enlaza.
- Ítems *<item>*: Usando estos elementos es posible describir los diferentes recursos que forman el canal definiendo su título, la URL donde es posible localizarlos, y una breve descripción o resumen de su contenido.
- Entrada de texto *<textinput>*: La definición de este elemento permite establecer mecanismos de envío de información a una URL determinada para realizar tareas de búsqueda de información o suscripción al canal. Para ello se indica un título que especifica el texto que aparecerá en el botón de ejecución, una breve descripción del método, el nombre del objeto de texto, y una URL donde se localiza el script CGI que procesa la tarea. Su uso es opcional.

La mayoría de los sub-elementos definidos en los elementos RSS están pensados para su conversión en elementos HTML en el momento de ser presentados a los usuarios.

Al construirse sobre RSS 0.9, está garantizada la compatibilidad “*hacia atrás*” ya que los procesadores de RSS 1.0 obvian aquellos elementos no soportados por

---

RSS 0.9, y aquellos que no interfieran en la estructura básica de un documento RSS 0.9 válido.

El vocabulario define una serie de restricciones sintácticas. Por ejemplo, no está permitido repetir sub-elementos definidos en el tercer nivel de profundidad del árbol jerárquico del canal. También se hacen algunas recomendaciones sobre la longitud máxima de la cadena de caracteres que cada elemento puede tomar como valor.

#### 3.4.2.4.2 MÓDULOS RSS

Además del vocabulario básico RSS dispone de una serie de módulos que lo extienden, permitiendo de esta forma añadir nuevas herramientas descriptivas de utilidad para resolver problemas específicos.

En concreto se definen un total de 25 módulos (que aparecen listados en la tabla X) de los cuales sólo tres de ellos, los módulos *DublinCore*, *Content* y *Syndication*, son módulos estándar u oficiales. El resto son considerados como módulos propuestos (en espera de aprobación definitiva).

<b>DublinCore</b>	Permite incluir en la descripción de canal y los ítems los elementos metadata definidos en el vocabulario Dublin Core.
<b>Content</b>	Se usa para describir el contenido de sitios web, con la posibilidad de indicar su codificación.
<b>Syndication</b>	Incluye información que indica a los agregadores el periodo de actualización del canal.
<b>Admin</b>	Permite definir elementos que faciliten determinados trámites administrativos como la comunicación de errores técnicos o datos estadísticos, generalmente vía correo electrónico.

<b>Aggregation</b>	Proporciona información suplementaria sobre la fuente de un ítem determinado para facilitar a los agregadores RSS la tarea de discriminar entre fuentes.
<b>Annotation</b>	Permite enriquecer la descripción de ítems, especificar su origen o referenciar otros recursos relacionados.
<b>Audio</b>	Se utiliza para la difusión de documentos de audio en formato mp3 utilizando como elementos las etiquetas definidas en el vocabulario de descripción ID3.
<b>Cc</b>	Este módulo permite especificar información relacionada con los derechos de autor que afectan a un ítem determinado.
<b>ChangedPage</b>	Permite identificar el servidor que notifica cuándo se actualiza el canal RSS.
<b>Company</b>	Define elementos específicos para enriquecer la descripción de compañías y empresas.
<b>Context</b>	Ofrece elementos para puntualizar información contextual sobre el origen del canal, sobre el solicitante, y sobre dónde y cómo debe ser procesado el canal.
<b>Dcterms</b>	Permite el uso de elementos cualificados de <i>Dublin Core</i> , es decir, aquellos elementos del vocabulario que extienden, complementan o refinan la expresividad de algunos los elementos del núcleo del vocabulario.
<b>Email</b>	Especifica elementos para describir encabezamientos de mensajes de correo electrónico.
<b>Event</b>	Permite la descripción precisa de eventos especificando fecha, lugar, tipología, y organizadores del mismo.
<b>Link</b>	De manera similar al mecanismo <i>link</i> de HTML 4.0, este módulo permite, entre otras cosas, asociar el canal o los ítems con diferentes versiones de los mismos (como, por ejemplo, una versión imprimible).

<b>Prism</b>	Permite definir elementos del vocabulario <i>Publishing Requirements for Industry Standard Metadata</i> (PRISM), específico para la descripción, reutilización e intercambio de publicaciones seriadas.
<b>Richequiv</b>	Este módulo permite el uso elementos XML como valor de propiedades definidas con elementos RSS 1.0.
<b>RSS091</b>	Proporciona elementos que permiten la compatibilidad del canal con agregadores RSS 0.91.
<b>Search</b>	El módulo permite enriquecer los ítems obtenidos como resultado de un proceso de búsqueda en la web añadiendo detalles como, por ejemplo, su relevancia.
<b>Servicestatus</b>	Extiende RSS añadiendo al vocabulario elementos que permiten describir el estado de los servidores y el servicio RSS.
<b>Slash</b>	Incluye elementos que permiten la descripción de ítems obtenidos de sitios web basados en la web de noticias Slashdot [ <a href="http://www.slashdot.org">http://www.slashdot.org</a> ].
<b>Streaming</b>	Ofrece capacidades para la descripción a nivel de ítems de ficheros multimedia de difusión simultánea ( <i>streaming media</i> ).
<b>Subscription</b>	Este módulo define elementos que añaden información sobre la suscripción al canal RSS.
<b>Taxonomy</b>	Permite identificar una taxonomía en la que están definidos los valores usados para especificar la temática del canal y sus ítems.
<b>Threading</b>	Define relaciones jerárquicas padre/hijo entre los ítems del canal.

Tabla 2. Módulos RSS 1.0



### 3.4.2.5 OWL (WEB ONTOLOGY LANGUAGE)

Como hemos visto, uno de los pilares básicos de la Web Semántica son las ontologías. Para definir las existen diferentes vocabularios que permiten describir un dominio específico definiendo los elementos (conceptos) que lo forman y las relaciones semánticas más o menos complejas que establecen entre ellos. El vocabulario que el W3C propone como estándar desde el año 2004 para definir ontologías es OWL (*Web Ontology Language*), un lenguaje que se define como extensión semántica de RDF Schema del que aprovecha la semántica definida para clases y propiedades, pero al que añade nuevos constructores más potentes y expresivos que permiten superar algunas de sus limitaciones expresivas.

OWL ofrece, por lo tanto, medios para especificar de una manera más precisa la semántica formal de un ámbito determinado de la realidad, y hace posible razonar con las ontologías para inferir conocimiento sobre diferentes aspectos de los elementos descritos. De esta manera, es posible determinar: (Antinou, van Harmelen 2004A):

- *Pertenencia a clases de instancias*: Si  $x$  es una instancia de la clase  $C$ , y  $C$  es una subclase de  $D$ , entonces  $x$  es una instancia de la clase  $D$ .
  - *Equivalencia de clases*: Si  $A$  es una clase equivalente a la clase  $B$ , y  $B$  es equivalente a la clase  $C$ , entonces  $A$  es equivalente a  $C$ .
  - *Consistencia del conocimiento*: Permite determinar errores lógicos en la definición de clases e instancias. Por ejemplo, si declaramos una instancia  $x$  que pertenece a la clase  $A$ , y  $A$  es a su vez subclase de la clase  $B \cap C$  y de la clase  $D$  (siendo  $D$  y  $B$  clases disjuntas), entramos en contradicción ya que  $A$  debería ser una clase vacía.
  - *Clasificación de instancias*: La imposición de determinados pares *propiedad/valor* como condición suficiente para ser miembro de una
-

clase, nos permite definir tantas instancias como sujetos cumplan esa condición.

#### 3.4.2.5.1 SINTAXIS BÁSICA

OWL, como vocabulario que se circunscribe dentro de la capa ontológica del modelo de Web Semántica, aprovecha la infraestructura definida en las capas inferiores. Así, de la capa sintáctica OWL utiliza la sintaxis XML para representar las ontologías, aunque también es frecuente recurrir a otras sintaxis simplificadas (como la que se utiliza en la propia especificación de OWL -(McGuinness, van Harmelen 2004)-), o incluso una sintaxis basada en el lenguaje de modelado unificado UML (*Unified Modeling Language*) [169] que permite representar las ontologías como si fueran diagramas de clases.

De la capa semántica hereda el modelo de datos RDF junto a gran parte de sus constructores (entendiendo por constructores cualquier primitiva, ya sea clase o propiedad) como *rdf:resource*, *rdf:property*, *rdf:about*, *rdf:ID*, *rdf:parseType* o *rdf:dataType*, que definen el entramado básico de funciones sobre el que se construye la Web Semántica. Como extensión semántica de RDF Schema, OWL hereda también hereda parte de sus primitivas, como las que permiten definir jerarquías dentro de clases y propiedades (*rdfs:subClassOf*, *rdfs:subPropertyOf*), o las que permiten definir el rango y el dominio de propiedades (*rdfs:range*, *rdfs:domain*).

Con las primitivas de RDF Schema *rdfs:label*, *rdfs:comment*, *rdfs:seeAlso*, *rdfs:isDefinedBy* (además del constructor propio *owl:versionInfo*) también es posible anotar recursos. Las anotaciones se definen bajo el constructor *owl:AnnotationProperty* y permiten añadir a la información representada en OWL modificadores cuyo objetivo es exclusivamente facilitar al usuario la comprensión del significado de un recurso específico; es decir, la inclusión de anotaciones en un recurso no mejoran su procesamiento por máquina.

---

Las ontologías definidas en OWL por lo general poseen una estructura bien definida que consta de una cabecera, una serie de axiomas (o definiciones) de clase, axiomas sobre propiedades y hechos sobre individuos. A continuación describimos cada uno de los componentes.

#### **A) CABECERA**

El primero de los componentes de una ontología OWL es una cabecera en la que se incluye información sobre la semántica y la compatibilidad de la propia ontología. La forman diferentes elementos que se definen como instancias de la clase *owl:ontologyProperty* y se declaran en el ámbito del constructor *owl:Ontology*.

Así, dentro de la cabecera es posible declarar la importación de clases y propiedades incluidas en otras ontologías externas a la nuestra utilizando el constructor *owl:imports* y referenciándolas debidamente mediante su correspondiente URI. De esta manera el significado de las clases y propiedades importadas se considera parte del significado de la ontología que las importa.

También es posible incluir en la cabecera información sobre las distintas versiones que existen de la ontología (la actual y anteriores) y sobre la compatibilidad entre ellas. Para ello se utiliza la siguiente serie de constructores:

- *owl:priorVersion* permite indicar mediante referencias URI si existen versiones anteriores de la ontología actual.
  - *owl:versionInfo* permite definir información (en formato texto, no procesable por máquina) sobre la versión actual de la ontología.
  - *owl:backwardCompatibilityWith* se utiliza para indicar versiones anteriores de una ontología con las que la actual es totalmente compatible. Es decir, estamos indicando al razonador que los elementos definidos en la versión previa de la ontología que indicamos se
-

interpretan de igual manera en la versión actual. No obstante, es posible que en la versión actual algunas clases y propiedades se mantengan exclusivamente por cuestiones de retro-compatibilidad y no tengan un uso real dentro de la ontología. En este caso, estas clases y propiedades obsoletas (que incluso pueden ser eliminadas en versiones posteriores) se definen mediante los constructores *owl:DeprecatedClass* y *owl:DeprecatedProperty* respectivamente, donde se indicará la relación que mantienen con la nueva clase o propiedad que las sustituye y cuyo uso es preferente.

- *owl:incompatibleWith* define la incompatibilidad de la ontología actual con la definida en este elemento.

## **B) ELEMENTOS DE CLASE**

Todas las clases posibles se definen como subclases de la clase genérica *owl:Thing* (incluida la clase *owl:Nothing* que representa el conjunto vacío). La manera más común de declarar clases en una ontología OWL consiste en asignar un nombre y un identificador único (una URI) a la clase dentro del ámbito del constructor *owl:Class*. No obstante, OWL dispone de otros cuatro mecanismos adicionales para declarar clases:

- Mediante la enumeración exhaustiva de los individuos que forman la clase.
- Mediante restricción de propiedades.
- Mediante la intersección o la unión de dos o más clases.
- Definiendo una clase como complemento lógico de otra.

Independientemente del mecanismo elegido para declararlas, es necesario utilizar clases anónimas dentro del ámbito de un constructor *owl:Class*. Partiendo de esta premisa, la manera menos compleja (aunque, por otro lado,

---

más laboriosa) consiste en la definición de la clase a partir de la enumeración de cada uno de los individuos que la forman. Estos deben estar identificados por su URI correspondiente y definidos dentro del ámbito de un constructor *owl:oneOf*.

Otra forma de crear nuevas clases es mediante la selección de individuos o elementos a partir de la restricción de los valores que puede tomar una propiedad determinada. Para ello es necesario utilizar la clase *owl:Restriction* dentro de la cual, a su vez, se define una propiedad *owl:onProperty* que permite identificar la propiedad sobre la que se va a realizar la restricción de valores. La restricción de valores se puede llevar a cabo de tres maneras diferentes: seleccionando todos los elementos de una enumeración de valores (*owl:allValuesFrom*), solo algunos de ellos (*owl:someValuesFrom*), o un único valor específico (*owl:hasValue*).

Otro de los mecanismos de los que dispone OWL para declarar clases consiste en la combinación de dos o más clases mediante el uso de operadores booleanos. De esta manera es posible combinar varias extensiones de clase con el constructor *owl:unionOf* o mediante su intersección utilizando *owl:intersectionOf*.

Por último, también es posible declarar nuevas clases a partir de la definición del complemento lógico de otras, utilizando para ello el constructor *owl:complementOf*.

Adicionalmente, en el ámbito de cualquier clase es posible declarar tanto jerarquías de clases, mediante el constructor *owl:subclass*, como clases equivalentes (*owl:equivalentClass*) y disjuntas (*owl:disjointWith*).

---

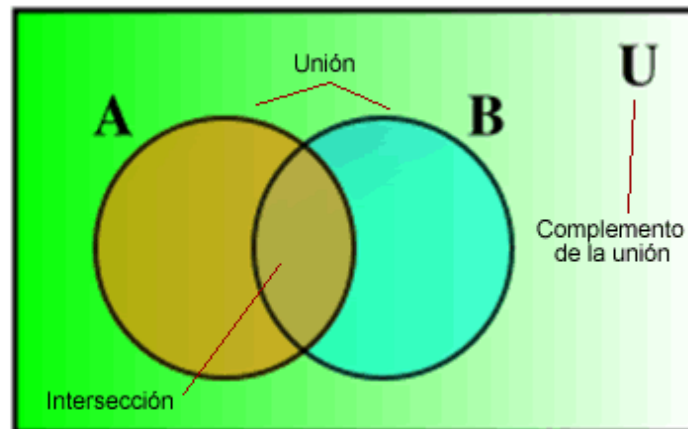


Fig.23 Unión, intersección y complemento de conjuntos

### C) AXIOMAS SOBRE PROPIEDADES

Otro de los elementos esenciales y necesarios dentro de la estructura de una ontología OWL son las propiedades. Al igual que en RDF Schema, en OWL se distinguen básicamente dos tipos de propiedades: propiedades de objeto (*owl:subjectProperty*), que son las que permiten relacionar unos individuos con otros, y propiedades de tipo de dato (*owl:DatatypeProperty*), que relacionan individuos con un literal determinado.

En el caso de las propiedades de tipo de dato OWL permite definir un rango de tipos de dato válidos para una propiedad determinada (declarados dentro de un constructor *owl:DataRange*). No obstante OWL, soporta sólo algunos de los tipos de dato nativos más comunes de XML Schema (como *string*, o *integer*) y no tipos de dato complejos construidos por los usuarios.

OWL dispone también de constructores para definir la cardinalidad de cualquier propiedad (es decir, el número de valores distintos que puede tomar esa propiedad concreta). En el caso de que la cardinalidad sea puntual (un único valor entero positivo) se utiliza *owl:cardinality*; si por el contrario la cardinalidad se mueve en un rango definido de valores podemos recurrir al uso de los constructores *owl:minCardinality* y *owl:maxCardinality* para definir las cardinalidades mínima y máxima, respectivamente.

Por último, y al igual que en el caso de las clases, también se pueden definir jerarquías de propiedades utilizando el constructor *owl:subProperty*, además de propiedades inversas (como la relación que pueden establecer comprador y vendedor) con el constructor *owl:inverseOf*, y propiedades equivalentes (*owl:equivalentProperty*).

#### D) PROPIEDADES ESPECIALES

Aparte de estas propiedades genéricas, OWL define otra serie de propiedades especiales que enriquecen su expresividad y permiten establecer relaciones más complejas entre los individuos descritos en una ontología. En concreto se distinguen cuatro tipos de propiedades especiales:

- Las propiedades transitivas (*owl:TransitiveProperty*), por lo general, permiten especificar relaciones de gradación entre elementos, como por ejemplo la propiedad “*ser mayor que*”. Así, si *a* es mayor que *b*, y *b* es mayor que *c*, entonces *a* es mayor que *c*.

$$\forall a, b, c \in A, aRb \wedge bRc \Rightarrow aRc$$

- Propiedades simétricas (*owl:SymmetricProperty*) son aquellas en las que si el par (*x,y*) es una instancia de la propiedad *P*, entonces el par (*y,x*) también lo es. Es el caso, por ejemplo, de la propiedad “*ser pariente de*”.

$$\forall x, y \in A, xRy \Rightarrow yRx$$

- Las propiedades funcionales (*owl:FunctionalProperty*) son una subclase especial de la clase *rdf:Property* que permiten definir como máximo un valor para cada objeto o individuo, como por ejemplo la propiedad “*edad*”. No se imponen restricciones sobre la unicidad del valor, por lo que más de un individuo puede tomar el mismo valor para esa propiedad.
-

- Propiedades inversas funcionales (*owl:InverseFunctionalProperty*) son las que definen un valor unívoco a cada instancia que no puede ser asignado a dos instancias diferentes. Es el caso de las propiedades que asignan identificadores o números personales como el ISBN a un libro o el DNI a un sujeto.

#### E) HECHOS SOBRE INSTANCIAS

Para definir instancias en una ontología OWL es necesario o bien incluir su descripción en formato RDF, o hacer referencia expresa a su ID (URI) cuando la definición es externa a la ontología. Utilicemos el mecanismo que utilizamos es conveniente diferenciarlas entre sí para evitar posibles inconsistencias semánticas y así facilitar en lo posible la tarea de agentes y razonadores semánticos.

Cuando queremos indicar que varios individuos son el mismo (aunque tengan URI's diferentes), disponemos del constructor *owl:sameAs*. Esta situación es bastante común, sobre todo cuando se agrega información de diferentes ontologías, ya que en estos casos un identificador único no garantiza la unicidad de una instancia (es posible importar individuos que aún siendo iguales han sido identificados con una URI distinta a la que nosotros hemos asignado). Utilizando este constructor evitamos cualquier tipo de ambigüedad.

Si por el contrario lo que queremos es diferenciar entre dos (o más) individuos utilizaremos la propiedad *owl:differentFrom*. No obstante, OWL también dispone de herramientas más elegantes para diferenciar entre colecciones de individuos mediante la clase *owl:AllDifferent*. Dentro de ella se define la propiedad *owl:distinctMembers* y en su ámbito se relacionan todas las instancias que son diferentes entre sí (referenciadas convenientemente por su correspondiente URI).

---



**F) ESPECIES DE OWL**

Como acabamos de ver, OWL es un lenguaje sólidamente construido sobre RDF y RDF Schema, que los amplía semánticamente y que garantiza la interoperabilidad entre capas sin imponer ningún tipo de restricciones sobre las posibilidades de inferencia y razonamiento. No obstante, esta gran expresividad tiene como contrapartida la imposibilidad de procesar e implementar de una manera eficiente las ontologías que están definidas utilizando todas las primitivas de OWL. Para paliar este problema se han definido tres especies o sub-lenguajes OWL que presentan diferentes niveles de expresividad (McGuinness, van Harmelen 2004):

b) OWL Full es la especie que integra todas las primitivas definidas en el lenguaje (para ser rigurosos ni siquiera debería considerarse como un sub-lenguaje) y permite combinarlas de forma arbitraria con RDF y RDF Schema.

- OWL DL (*Description Logics*) por su parte restringe el modo en que se pueden utilizar los constructores de OWL y RDF de manera que las ontologías se ajustan a los requerimientos de los razonadores basados en lógica descriptiva.
- OWL Lite impone condiciones más estrictas sobre los constructores admitidos para OWL DL, lo cual disminuye su capacidad expresiva pero facilita el razonamiento con las ontologías diseñadas con este sub-lenguaje.

No obstante, las tres especies no son lenguajes independientes, sino que están interrelacionados a través de una serie de reglas de compatibilidad definidas en la especificación:

- Toda ontología OWL Lite es una ontología OWL DL válida.
  - Toda ontología OWL DL es una ontología OWL Full válida.
-

- Toda inferencia realizada con una ontología OWL Lite es una conclusión válida en OWL DL.
- Toda inferencia realizada con una ontología OWL DL es una conclusión válida en OWL Full.
- Todas las variedades de OWL usan el modelo de datos RDF.
- Las instancias se declaran usando descripciones RDF.
- Las primitivas básicas de OWL *owl:Class*, *owl:DataTypeProperty* y *owl:ObjectProperty* se definen como especializaciones de los constructores RDF Schema análogos.

Atendiendo a estos requisitos de partida se establecen los diferentes elementos de OWL que pueden ser utilizados en cada una de las especies.

En OWL Full no existe ningún tipo de restricción y por lo tanto se pueden utilizar todos los constructores definidos en la especificación para diseñar una ontología. Además, la equivalencia de determinados constructores OWL Full con constructores RDF Schema es absoluta. Este es el caso, por ejemplo, de *owl:Class* y *rdfs:Class*, equivalentes en OWL Full pero no así en OWL DL y OWL Lite donde *owl:Class* se considera una subclase de *rdfs:Class*. Como hemos comentado con anterioridad, esta característica hace que el uso de OWL Full permita al desarrollador explotar al máximo las capacidades de metamodelado de RDF y garantizar una completa compatibilidad con las capas inferiores (es decir, una conclusión obtenida con una ontología OWL Full podría ser válida para un esquema RDF), pero a costa de perder garantías sobre la eficiencia del procesamiento de las ontologías definidas con OWL Full por parte de los motores de razonamiento.

El segundo sub-lenguaje, OWL DL, se define como una partición del OWL Full al que se le aplican una serie de restricciones:

---

- Cualquier elemento definido en OWL DL puede pertenecer exclusivamente a una categoría de entre los diferentes elementos (clases, propiedades, tipo de dato, etc.) que se pueden definir en RDF Schema. Esto implica que, por ejemplo, una clase no pueda ser considerada al mismo tiempo un individuo. Además, la pertenencia a esta categoría debe ser expresada convenientemente de forma explícita (es decir, no se podría inferir de forma directa que B es una subclase de A sin antes indicar que A y B son clases).
- Las propiedades de objetos y de tipo de dato son consideradas como elementos disjuntos, por lo que, por ejemplo, para las propiedades de tipo de dato no será posible definir propiedades inversas, inversas funcionales, transitivas o simétricas.
- Tampoco está permitido definir restricciones de cardinalidad sobre las propiedades (o subpropiedades) transitivas o sus inversas.
- Las clases anónimas solo se pueden definir en el rango y dominio de *owl:equivalentClass* y *owl:disjointWith*, y en el rango de *rdfs:subClassOf*. El resto de clases deben estar identificadas por su correspondiente ID.
- El objeto de las anotaciones debe ser un literal, una URI o un individuo (de esta manera en las anotaciones no sería posible, por ejemplo, definir propiedades de rango o dominio).

De esta manera las ontologías OWL DL se ajustan a los requerimientos definidos por los sistemas de razonamiento basados en lógica descriptiva y suponen para los desarrolladores una herramienta semántica plenamente operativa (aunque el usar una versión menos expresiva de OWL suponga una merma de la interoperabilidad con las capas inferiores del modelo de Web Semántica).

El último sub-lenguaje es OWL Lite, la versión expresivamente más limitada de las tres, que añade una serie de restricciones a las ya impuestas en OWL DL:

---

- No está permitido usar los constructores *owl:oneOf*, *owl:disjointWith*, *owl:unionOf*, *owl:complementOf*, *owl:DataRange* y *owl:hasValue*.
- La cardinalidad solo puede tomar los valores 0 o 1, ya sea para definir rangos de cardinalidad o para cardinalidades exactas.
- Los constructores *owl:equivalentClass*, *owl:subclassOf* y *owl:intersectionOf* solo pueden ser utilizados con clases identificadas o restricciones. Los constructores *owl:allValuesFrom*, *owl:someValuesFrom* necesitan definir sus objetos con clases o tipos de datos identificados. Por su parte *rdf:type*, *rdfs:Domain* y *rdfs:range* solo puede ser utilizados cuando su objeto sean clases identificadas, clases identificadas o restricciones, y clases o tipos de datos identificados respectivamente.

Esta versión es mucho más simple y permite el desarrollo de ontologías con las que es más fácil operar (ya que disponen de capacidad más que suficiente para trabajar en determinados ámbitos donde los requerimientos no son demasiado exigentes), pero tiene el inconveniente de que carece de gran parte de la interoperabilidad que ofrecen las otras dos especies.

Como vemos, la elección de una u otra especie de OWL es cuestión únicamente de determinar qué problema queremos resolver y cuáles son las necesidades específicas del sistema en el que queremos implantar una ontología determinada (o lo que es lo mismo, definir en qué grado deseamos preservar la interoperabilidad o facilitar el razonamiento de una ontología).

---

### 3.4.3 MODELADO LINGÜÍSTICO DIFUSO

En este apartado vamos a analizar las técnicas lingüísticas para el manejo de información sobre las que vamos a basar nuestra investigación y que, aplicadas conjuntamente con las tecnologías de Web Semántica, van a proporcionar una mayor flexibilidad en el proceso de tratamiento de la información, especialmente en los casos en los que se produce una interacción con los usuarios.

#### 3.4.3.1 INTRODUCCIÓN

La forma en que podemos evaluar la información relativa a un problema específico depende de varios factores. En muchas ocasiones la información puede ser evaluada de manera cuantitativa fruto de una medición empírica, como cuando se determina la altura o el peso de un individuo mediante un valor numérico preciso asociado a una unidad física de medida.

Sin embargo, hay situaciones en las que trabajar con información precisa de manera cuantitativa simplemente no es posible, bien porque no están disponibles los elementos necesarios para llevar a cabo una medición exacta de esa información, o bien porque nos basta con la determinación de un valor aproximado (como por ejemplo, en la evaluación de aspectos relacionados con la percepción subjetiva de un individuo). Ante casos de esta naturaleza la información puede ser evaluada cualitativamente mediante el uso de términos lingüísticos, como cuando realizamos la valoración sobre una película que hemos visto utilizando expresiones como *buena*, *regular* o *mala*.

Para gestionar eficientemente este tipo de problemas resulta adecuado aplicar técnicas de modelado lingüístico difuso, un enfoque aproximado basado en la Teoría de Conjuntos Difusos que permite representar aspectos cualitativos en la forma de valores lingüísticos que se definen mediante lo que se conoce como variables lingüísticas (Zadeh 1975). Estas variables se caracterizan porque los valores que pueden tomar no son números, sino palabras o sentencias expresadas en lenguaje natural o artificial. Cada valor lingüístico se compone de

---

un valor o etiqueta y un valor semántico o significado. La etiqueta es una palabra o sentencia perteneciente a un conjunto de términos lingüísticos y el significado es un subconjunto difuso en un universo de discurso.

Son numerosos los ejemplos en los que el modelado lingüístico difuso ha demostrado ser una herramienta muy útil para gestionar problemas que se presentan, por ejemplo, en la toma de decisiones (Herrera, Herrera-Viedma, Verdegay 1996), en la evaluación de la calidad de la información de documentos web (Herrera-Viedma, Peis 2003), o en los modelos de recuperación de información (Herrera-Viedma 2001).

En este apartado vamos a hacer un breve repaso a los diferentes enfoques del modelado lingüístico difuso hallados en la literatura, centrándonos especialmente en el modelado lingüístico difuso ordinal que es el que utilizamos en el modelo que presentamos en este trabajo.

#### **3.4.3.2 MODELADO LINGÜÍSTICO DIFUSO CLÁSICO**

En el mundo real hay determinadas ocasiones en las que para evaluar información relativa a un problema específico es suficiente con recurrir a la lógica tradicional. Esta define una semántica muy simple donde es posible establecer si un aserto es (o no) verdadero o falso en la estructura del mundo. En otras palabras, la lógica tradicional es bivaluada y solo es posible definir dos grados de verdad (por ejemplo, en condiciones normales la luz de una habitación puede estar encendida o apagada).

No obstante, esta aproximación no es la adecuada en otras muchas ocasiones en las que es posible distinguir diferentes rangos de valoración (o grados de verdad) y los valores pueden tener distinta naturaleza.

Ante estos casos, la teoría de Conjuntos Difusos dispone de los medios para poder caracterizar de una manera progresiva estos diferentes grados de verdad haciendo que la transición de unos a otros sea suave. Esencialmente esta teoría

---

se centra en modelar aquellos problemas donde los enfoques clásicos de la Teoría de Conjuntos y la Teoría de la Probabilidad no resultan adecuados. Se basa en la generalización de la noción clásica de conjunto e introduce el concepto de ambigüedad, de manera que los conjuntos difusos proporcionan una nueva manera de representar la imprecisión e incertidumbre presentes en determinados problemas.

Para enmarcar adecuadamente esta teoría hagamos en primer lugar una revisión de los principios básicos de la Teoría de Conjuntos Difusos que van a ser utilizados, no solo en el modelado lingüístico difuso clásico, sino también en el resto de tipos de modelado lingüístico.

#### **3.4.3.2.1 CONJUNTOS DIFUSOS Y FUNCIONES DE PERTENENCIA**

La noción de conjunto refleja la tendencia a organizar, generalizar y clasificar el conocimiento sobre los objetos del mundo real. El encapsulamiento de los objetos es una colección cuyos miembros comparten una serie de características o propiedades que implican la noción de conjunto.

Los conjuntos introducen una noción de dicotomía, que en esencia es una clasificación binaria excluyente que implica la aceptación o el rechazo de la pertenencia de un objeto a una categoría determinada. Habitualmente la pertenencia se representa mediante el valor 1 y la no pertenencia con el valor 0. Esta decisión de aceptar o rechazar se expresa mediante una función característica, según las propiedades que posean los objetos del conjunto.

El concepto de conjunto difuso rebaja la exigencia de este requerimiento y admite valores intermedios en la función característica que se denomina función de pertenencia. Esto permite una interpretación más realista de la información, puesto que la mayoría de las categorías que describen los objetos del mundo real, no tienen unos límites claros y bien definidos.

---

Un conjunto difuso puede definirse como una colección de objetos con valores de pertenencia entre 0 (exclusión total) y 1 (pertenencia total). Los valores de pertenencia expresan los grados con los que cada objeto es compatible con las propiedades o características que definen el conjunto.

Consideremos por ejemplo, el concepto “*temperatura*”, en un contexto donde esta oscila entre 0°C y 100°C. Si definimos el conjunto difuso “*temperatura alta*”, una temperatura con un valor de 90°C tendría un grado de pertenencia de 1 a dicho conjunto, mientras que para el caso de una temperatura de 7°C dicho grado de pertenencia sería 0.

Formalmente, un conjunto difuso  $A$  definido sobre un dominio o universo de discurso  $U$  está caracterizado por una función de pertenencia que asocia a cada elemento del conjunto el grado con que pertenece a dicho conjunto en el intervalo  $[0,1]$ :

$$\mu_{\tilde{A}}: U \rightarrow [0, 1]$$

Así, un conjunto difuso  $A$  sobre  $U$  se puede representar como un conjunto de pares ordenados de un elemento perteneciente a  $U$  y su grado de pertenencia,

$$\tilde{A} = \{(x, \mu_{\tilde{A}}(x)) / x \in U, \mu_{\tilde{A}}(x) \in [0,1]\}$$

#### 3.4.3.2.2 VARIABLES LINGÜÍSTICAS

Una variable lingüística se caracteriza por un valor sintáctico o etiqueta que es una palabra o frase perteneciente a un conjunto de términos lingüísticos, y por un valor semántico o significado de dicha etiqueta que viene dado por un subconjunto difuso en un universo de discurso.

Formalmente, una variable lingüística (Zadeh 1975) está caracterizada por una 5-tupla  $(H, T(H), U, G, M)$ , donde  $H$  es el nombre de la variable,  $T(H)$  (o solo  $T$ ) simboliza el conjunto de términos lingüísticos de  $H$ , es decir, el conjunto de



nombres de valores lingüísticos de  $H$ , donde cada valor es una variable difusa denotada genéricamente como  $X$  y que varía según el universo de discurso  $U$ , que está asociado con una variable base denominada  $u$ ;  $G$  es una regla sintáctica (que normalmente toma forma de gramática) para generar los nombres de los valores  $H$ ; y  $M$  es una regla semántica para asociar significado a cada elemento  $H$ , que será un subconjunto difuso de  $U$ .

### 3.3.3.2.3 ELECCIÓN DEL CONJUNTO DE TÉRMINOS LINGÜÍSTICOS

Uno de los aspectos fundamentales que debemos de tener en cuenta es la elección del conjunto de términos lingüísticos con las semánticas asociadas que vamos a usar para aplicar en un problema determinado y así proporcionar a una fuente de información un número reducido de términos con los que poder expresarla.

Para ello debemos establecer la granularidad de la incertidumbre (Bonissone, Decker 1986), es decir, la cardinalidad del conjunto de términos lingüísticos usado para expresar y representar la información. La cardinalidad debe ser suficientemente baja como para no imponer una precisión excesiva en la información que se quiera expresar y suficientemente alta como para conseguir una discriminación de las valoraciones en un número limitado de grados.

Habitualmente la cardinalidad usada en los modelos lingüísticos suele ser un valor impar, como 7 o 9, no superando las 11 o 13 etiquetas. Estos valores clásicos de cardinalidad están basados en la línea de observación de Miller (1956) sobre la capacidad humana, en la que se indica que para un individuo el número aproximado de términos que puede manejar y recordar de una manera eficiente oscila entre los 7 y 9 términos. El término que ocupa la posición central representa una valoración aproximada de 0.5, y el resto de términos se sitúan simétricamente alrededor de este punto medio (Bonissone, Decker 1986).

Una vez establecida la cardinalidad del conjunto de términos lingüísticos, hay

---

que definir dicho conjunto, es decir, cuáles van a ser las etiquetas lingüísticas y su semántica asociada.

En el caso del modelado lingüístico difuso clásico se utiliza un enfoque basado en una gramática libre de contexto  $G$  (Bonissone, Decker 1986) (Zadeh 1975), donde el conjunto de términos pertenece al lenguaje generado por  $G$ . Una gramática generadora  $G$  es una 4-tupla  $(VN, VT, I, P)$ , siendo  $VN$  el conjunto de símbolos no terminales,  $VT$  el conjunto de símbolos terminales,  $I$  el símbolo inicial y  $P$  el conjunto de reglas de producción. La elección de estos cuatro elementos determinará la cardinalidad y forma del conjunto de términos lingüísticos. Entre los símbolos terminales y no terminales de  $G$  podemos encontrar términos primarios (como, por ejemplo, *alto, medio, bajo*), modificadores (por ejemplo, *no mucho, muy, más o menos*), relaciones (por ejemplo, *mayor que, menor que*) y conectivos (por ejemplo, *y, o, pero*).

Siendo  $I$  cualquier término primario y usando las reglas definidas en  $P$ , construimos el conjunto de términos lingüísticos  $H = \{muy\ alto, alto, medio, \dots\}$ . La semántica del conjunto de términos lingüísticos se define utilizando números difusos en el intervalo  $[0, 1]$ , donde cada número difuso es descrito por una función de pertenencia basada en ciertos parámetros o reglas semánticas.

Dentro de este enfoque es posible definir operadores de agregación de información lingüística, que básicamente extienden las operaciones de la lógica tradicional para aplicarlas sobre las funciones de pertenencia. No obstante, presentan el inconveniente de que el resultado es otro conjunto difuso que no se corresponde con ninguna etiqueta del conjunto de términos original.

Para ceñirnos a este conjunto de términos es necesario llevar a cabo un proceso de aproximación lingüística que permita escoger aquella etiqueta del conjunto original cuyo significado sea el más similar posible (de acuerdo a alguna métrica) a la semántica del conjunto difuso no etiquetado obtenido como resultado de la operación de agregación.

### 3.4.3.3 MODELADO LINGÜÍSTICO DIFUSO ORDINAL

Este enfoque (Herrera, Herrera-Viedma, Verdegay 1996) (Herrera, Herrera-Viedma 1997) está basado en la definición de términos primarios con una estructura ordenada (Delgado et al. 1998), y se presenta como una alternativa para reducir la complejidad del enfoque tradicional donde es necesario establecer una gramática para definir la semántica del conjunto de términos lingüísticos. Consiste en definir directamente un conjunto de términos, considerándolos todos primarios y distribuidos sobre una escala en la que hay definida una relación de orden total.

Un enfoque lingüístico difuso ordinal se define considerando un conjunto de etiquetas finito y totalmente ordenado:

$$S = \{s_i\}, i \in \{0, \dots, g\} \text{ con } s_i \geq s_j \text{ si } i \geq j$$

y con una cardinalidad impar (por lo general 7 o 9 etiquetas).

La semántica del conjunto de etiquetas es establecida según la estructura ordenada del conjunto de etiquetas, considerando que cada etiqueta del par  $(s_i, s_{g-1})$  es igualmente informativa. Por ejemplo, podríamos usar el siguiente conjunto de 7 etiquetas para representar la información lingüística:

$$S = \{s_0 = N, s_1 = MB, s_2 = B, s_3 = M, s_4 = A, s_5 = MA, s_6 = T\}$$

donde  $sa < sb$  si y sólo si  $a < b$ , y donde  $N=Nulo$ ,  $MB=Muy bajo$ ,  $B=Bajo$ ,  $M=Medio$ ,  $A=Alto$ ,  $MA=Muy alto$ , y  $T=Total$ .

Además, podemos asociar con cada término lingüístico un número difuso definido en el intervalo  $[0, 1]$ .

---

Una forma de caracterizar un número difuso es usar una representación basada en parámetros de su función de pertenencia. Puesto que las valoraciones lingüísticas dadas por los usuarios son únicamente aproximaciones, algunos autores consideran que las funciones de pertenencia trapezoidales lineales son suficientemente buenas para capturar la imprecisión.

La representación paramétrica es obtenida generalmente mediante funciones de pertenencia triangulares que se representan a partir de una 3-tupla  $(a, b, c)$ , donde  $a$  indica el punto donde el valor de pertenencia vale 1, y donde  $b$  y  $c$  indican los límites izquierdo y derecho respectivamente. Como ejemplo podemos considerar el siguiente conjunto de 7 términos (Fig. 24) :

$$s_0 (N= \text{Nulo})= (0, 0, 0.17)$$

$$s_1 (MB= \text{Muy bajo})= (0.17, 0, 0.33)$$

$$s_2 (B= \text{Bajo})= (0.33, 0.17, 0.5)$$

$$s_3 (M= \text{Medio})= (0.5, 0.33, 0.67)$$

$$s_4 (A= \text{Alto})= (0.67, 0.5, 0.83)$$

$$s_5 (MA= \text{Muy alto})= (0.83, 0.67, 1)$$

$$s_6 (T= \text{Total})= (1, 0.83, 1).$$

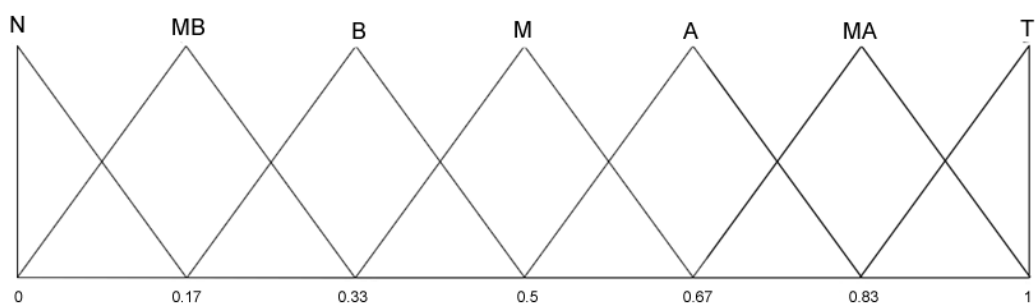


Fig.24 Semántica asociada a un conjunto de 7 términos lingüísticos.

**A) OPERADORES DE AGREGACIÓN**

En cualquier enfoque lingüístico necesitamos operadores para el manejo de la información lingüística. Una ventaja del enfoque lingüístico difuso ordinal es la simplicidad y agilidad de su modelo computacional. Está basado en el cálculo simbólico (Herrera, Herrera-Viedma, Verdegay 1996) (Herrera, Herrera-Viedma 1997) y actúa operando directamente sobre las etiquetas, teniendo en cuenta el orden de las valoraciones lingüísticas en la estructura ordenada de las etiquetas. Habitualmente, el modelo lingüístico difuso ordinal para la computación con palabras se define estableciendo:

1. Un operador de negación,
2. Operadores de comparación basados en la estructura ordenada de los términos lingüísticos
3. Operadores apropiados para la agregación de información lingüística difusa ordinal.

En la mayoría de los enfoques lingüísticos difusos ordinales, a partir de la semántica asociada a los términos lingüísticos el operador de negación se define como:

$$NEG(s_i) = s_j / j = g - i$$

También podemos definir dos operadores de comparación de términos lingüísticos:

- *Operador de maximización:*  $MAX(s_i, s_j) = s_i$  si  $s_i \geq s_j$
  - *Operador de minimización:*  $MIN(s_i, s_j) = s_i$  si  $s_i \leq s_j$
-

A partir de estos operadores es posible definir operadores automáticos y simbólicos de agregación de información lingüística, como por ejemplo el operador de agregación de información lingüística no ponderada LOWA ([30]) y el operador de información lingüística ponderada LWA (Herrera, Herrera-Viedma 1997). En este apartado nos vamos a centrar en exclusiva en el operador LOWA ya que es el que utilizamos en el modelo que presentamos en esta memoria.

### B) OPERADOR LOWA

Sea  $A = \{a_1, \dots, a_m\}$  un conjunto de etiquetas a agregar, con  $a_i \in S$ , entonces el operador LOWA,  $\phi$ , se define como:

$$\begin{aligned} \phi(a_1, \dots, a_m) &= W \cdot B = C^m \{\omega_k, b_k, k = 1, \dots, m\} = \\ &= \omega_1 \odot b_1 \oplus (1 - \omega_1) \odot C^{m-1} \{\beta_h, b_h, h = 2, \dots, m\} \end{aligned}$$

donde  $W = [\omega_1, \dots, \omega_m]$ , es un vector de ponderación, tal que:

$$\omega_i \in [0, 1] \quad \text{y} \quad \sum_{i=1}^m \omega_i = 1$$

y  $\beta_h = \omega_h / \sum_{k=2}^m \omega_k$ ,  $h = 2, \dots, m$ , siendo  $B = (b_1, \dots, b_m)$  un vector asociado a  $A$ , tal que:

$$B = \sigma(A) = (a_{\sigma(1)}, \dots, a_{\sigma(m)})$$

donde,  $a_{\sigma(j)} \leq a_{\sigma(i)} \quad \forall i \leq j$ , siendo  $\sigma$  una permutación definida sobre el conjunto de etiquetas  $A$ .  $C^m$  es el operador de combinación convexa de  $m$  etiquetas [26], de modo que si  $m = 2$ , entonces se define como:

$$C^2\{\omega_i, b_i, i = 1, 2\} = \omega_1 \odot s_j \oplus (1 - \omega_1) \odot s_i = s_k, \quad s_j, s_i \in S, (j \geq i),$$

$$k = \text{MIN} \{g, i + \text{round}(\omega_1 \cdot (j - i))\}$$

Donde *round* simboliza el operador de redondeo usual, y  $b_1 = s_j$ ,  $b_2 = s_i$ . Por otro lado,  $\omega_j = 1$ ,  $\omega_i = 0$  con  $i \neq j \forall i$ , entonces el operador de combinación se define como:

$$C^m\{\omega_i, b_i, i = 1, \dots, m\} = b_j$$

### C) EJEMPLO DE APLICACIÓN DEL OPERADOR LOWA

Supongamos  $m= 2$ ,  $W = [0.4, 0.6]$  y que usamos el conjunto de 7 etiquetas lingüísticas antes definido:

s0 (N= Nulo)= (0, 0, 0.17)

s1 (MB= Muy bajo)= (0.17, 0, 0.33)

s2 (B= Bajo)= (0.33, 0,17, 0.5)

s3 (M=Medio)= (0.5, 0.33, 0.67)

s4 (A= Alto)= (0.67, 0.5, 0.83)

s5 (MA= Muy alto)= (0.83, 0.67, 1)

s6 (T= Total)= (1, 0.83, 1).

Con estos valores se obtienen los siguientes resultados al aplicar el operador:

$$k_{46} = \text{MIN} \{6, 3 + \text{round} (0.4 * (5 - 3))\} = 4 \Rightarrow l_{k46} = \text{Alto}$$

	Nulo	Muy bajo	Bajo	Medio	Alto	Muy alto	Máximo
Nulo	Nulo	Muy bajo	Muy bajo	Muy bajo	Bajo	Bajo	Bajo
Muy Bajo		Muy bajo	Bajo	Bajo	Bajo	Medio	Medio
Bajo			Bajo	Bajo	Medio	Medio	Alto
Medio				Medio	Medio	Alto	Alto
Alto					Alto	Muy alto	Muy alto
Muy alto						Muy alto	Muy alto
Máximo							Máximo

Tabla 3. Resultados del operador LOWA con  $m=2$  y  $\omega_1=0.4$

### 3.4.3.4 MODELADO LINGÜÍSTICO BASADO EN 2-TUPLAS

El modelado lingüístico difuso basado en 2-tuplas (Herrera, Martínez 2000) es un tipo de modelado lingüístico difuso que permite reducir la pérdida de información, y por lo tanto de precisión, que habitualmente se produce en el modelado lingüístico difuso ordinal. La principal ventaja de este modelo, es que mantiene la sencillez de cálculo con palabras del enfoque lingüístico ordinal pero sin pérdida de información puesto que utiliza un modelo continuo de representación de la información en lugar de valores discretos. Para definirlo, tenemos que establecer el modelo de representación y el modelo computacional de las 2-tuplas para representar y agregar la información lingüística respectivamente.

#### A) MODELO DE REPRESENTACIÓN LINGÜÍSTICA BASADA EN 2-TUPLAS

Sea  $S = \{s_0, \dots, s_g\}$  un conjunto de términos lingüísticos con cardinalidad impar, donde el término intermedio representa una valoración de aproximadamente 0.5, y con el resto de términos del conjunto distribuidos simétricamente alrededor de ese punto intermedio. Asumimos que la semántica asociada con cada una de las etiquetas viene dada por medio de funciones de pertenencia triangulares, representadas por 3-tuplas  $(a, \alpha, \beta)$  y consideramos todos los términos distribuidos sobre una escala sobre la que hay establecida una relación de orden total, es decir,  $s_i \leq s_j \Leftrightarrow i \leq j$ .

En este contexto lingüístico difuso, si mediante un método simbólico de agregación de información lingüística (Herrera, Herrera-Viedma 1997),(Herrera, Herrera-Viedma, Verdegay 1996) obtenemos un valor  $\beta \in [0, g]$  y  $\beta \notin \{0, \dots, g\}$ , podemos usar una función de aproximación para expresar el resultado obtenido como un valor de  $S$ .



Un concepto fundamental de esta aproximación al modelado lingüístico es el de “*traslación simbólica*” (Herrera, Martínez 2000). Sea  $\beta$  el resultado de una agregación de los índices de un conjunto de etiquetas valoradas sobre un conjunto de términos lingüísticos  $S$ , es decir, el resultado de una operación de agregación simbólica,  $\beta \in [0, g]$ . Dados  $i = \text{round}(\beta)$  y  $\alpha = \beta - i$  dos valores, tales que,  $i \in [0, g]$  y  $\alpha \in [-0.5, 0.5)$  entonces  $\alpha$  es lo que denominamos “*traslación simbólica*”, que expresa la diferencia de información entre la información expresada por  $\beta$  y la etiqueta lingüística  $s_i$  más cercana a  $S$ .

El enfoque lingüístico difuso basado en 2-tuplas es desarrollado a partir del concepto de traslación simbólica, representando la información lingüística por medio de 2-tuplas  $(s_i, \alpha_i)$ ,  $s_i \in S$  y  $\alpha_i \in [-0.5, 0.5)$ , donde  $s_i$  representa la etiqueta lingüística, y  $\alpha_i$  es un valor que expresa la traslación de  $\beta$  al índice  $i$  de la etiqueta más cercana en el conjunto de términos lingüísticos ( $s_i \in S$ ).

Este modelo define un conjunto de funciones de transformación entre valores numéricos y 2-tuplas.

1. Sea  $s_i \in S$  un término lingüístico, su representación mediante una 2-tupla se obtiene mediante la función  $\theta$ :

$$\theta : [0, g] \rightarrow S \times [-0.5, 0.5)$$

$$\theta(s_i) = (s_i, 0) / s_i \in S$$

2. Sea  $S = \{s_0, \dots, s_g\}$  un conjunto de términos lingüísticos y  $\beta \in [0, g]$  un valor que representa el resultado de una operación de agregación simbólica, la 2-tupla que expresa la información equivalente a  $\beta$  se obtiene mediante la siguiente función:

$$\Delta : [0, g] \rightarrow S \times [-0.5, 0.5)$$

$$\Delta(\beta) = (s_i, \alpha), \text{ con } \begin{cases} s_i & i = \text{round}(\beta) \\ \alpha = \beta - i & \alpha \in [-0.5, 0.5) \end{cases}$$

donde *round* es el operador de redondeo usual,  $s_i$  es la etiqueta cuyo índice es el más cercano a  $\beta$  y  $\alpha$  es el valor de la traslación simbólica.

Así, por ejemplo, suponiendo el siguiente conjunto de términos lingüísticos  $S = \{s_0, s_1, s_2, s_3, s_4, s_5, s_6\}$  y que como resultado de una operación de agregación simbólica se obtiene el valor  $\beta = 3.3$ , la representación de este valor mediante una 2-tupla lingüística, sería  $\Delta(\beta) = (s_3, +0.3)$ .

3. Sea  $S = \{s_0, \dots, s_g\}$  un conjunto de términos lingüísticos y  $(s_i, \alpha)$  una 2-tupla. Se define la función  $\Delta^{-1}$ , tal que aplicada sobre una 2-tupla  $(s_i, \alpha)$  devuelve su valor numérico  $\beta \in [0, g]$ .

$$\Delta^{-1} : S \times [-0.5, 0.5) \rightarrow [0, g]$$

$$\Delta^{-1}(s_i, \alpha) = i + \alpha = \beta$$

## B) MODELO COMPUTACIONAL LINGÜÍSTICO DE LAS 2-TUPLAS

El modelo computacional que nos permite operar sobre la representación lingüística de las 2-tuplas, basándonos en los operadores de comparación, negación y agregación de 2-tuplas. A continuación pasamos a caracterizar cada uno de ellos:

### 1. Operador de comparación de 2-tuplas:

La comparación de información lingüística representada por medio de 2-tuplas se realiza de acuerdo a un orden lexicográfico común. Consideremos dos 2-tuplas  $(s_k, \alpha_1)$  y  $(s_l, \alpha_2)$  que representan cálculos de información:

- si  $k < l$  entonces  $(s_k, \alpha_1)$  es menor que  $(s_l, \alpha_2)$ .
- si  $k = l$  entonces
  - si  $\alpha_1 = \alpha_2$  entonces  $(s_k, \alpha_1)$  y  $(s_l, \alpha_2)$  representan la misma información,
  - si  $\alpha_1 < \alpha_2$  entonces  $(s_k, \alpha_1)$  es menor que  $(s_l, \alpha_2)$ ,
  - si  $\alpha_1 > \alpha_2$  entonces  $(s_k, \alpha_1)$  es mayor que  $(s_l, \alpha_2)$ .

### 2. Operador de negación de 2-tuplas:

El operador de negación sobre una 2-tupla se define como:

$$Neg((s_i, \alpha)) = \Delta(g - (\Delta^{-1}(s_i, \alpha)))$$

siendo  $g + 1$  la cardinalidad del conjunto de etiquetas  $S$ .

### 3. Operadores de agregación de 2-tuplas:

La agregación de información consiste en obtener un valor que resuma un conjunto de valores, por lo que el resultado de la agregación de un conjunto de 2-tuplas debe ser una 2-tupla. A lo largo de la literatura podemos encontrar numerosos operadores de agregación que nos permiten combinar la información de acuerdo a distintos criterios como los operadores media aritmética, media ponderada y media lingüística ponderada. Cualquiera de estos operadores ya existentes puede ser fácilmente extendido para trabajar con 2-tuplas, usando funciones  $\Delta$  y  $\Delta^{-1}$

---

que transforman valores numéricos en 2-tuplas y viceversa sin pérdida de información.

#### **3.4.3.5 OTRAS APROXIMACIONES AL MODELADO LINGÜÍSTICO DIFUSO**

Otros tipos de modelado lingüístico que se definen en la literatura y que pretenden mejorar el modelado lingüístico difuso clásico ofreciendo nuevas perspectivas y soluciones.

Así por ejemplo, el enfoque lingüístico difuso multi-granular permite trabajar con distintos conjuntos de etiquetas en aquellos casos en los que no sea eficiente valorar la información usando un mismo sistema de valores (Herrera-Viedma et al. 2005) (Herrera, Herrera-Viedma, Martínez 2000).

Por su parte, el modelado lingüístico difuso no balanceado (Herrera et al. 2003) permite manejar problemas en los que asumimos que no todos los términos del conjunto de etiquetas son igualmente informativos, es decir, las etiquetas no están distribuidas simétricamente (como ha ocurrido en los diferentes enfoques revisados hasta el momento). En este caso, es posible gestionar los conjuntos de términos lingüísticos con distintos niveles de discriminación a ambos lados del término medio.

Según Herrera y colaboradores (2003), el uso de este tipo de conjuntos de términos lingüísticos es adecuado para expresar los pesos de importancia en las consultas y representar los grados de relevancia de los documentos en sistemas de recuperación de información.

---



# MODELO DE SERVICIO SEMÁNTICO-DIFUSO DE DSI PARA BIBLIOTECAS DIGITALES

En este capítulo presentamos un modelo de servicio de difusión selectiva de información que pretende mejorar el acceso de los usuarios de las bibliotecas digitales a recursos de su interés combinando las tecnologías y técnicas analizadas en los capítulos previos, es decir, utilizando tecnologías de Web Semántica (para definir descripciones enriquecidas de recursos, esquemas de conceptos y herramientas de redifusión de contenidos), técnicas de modelado lingüístico difuso (tanto ordinal como basadas en 2-tuplas), así como técnicas de filtrado de información basadas en contenido y colaborativas.

### **4.1 INTRODUCCIÓN**

El problema de la recuperación de información en la Web es uno de los principales desafíos hoy día para la comunidad científica. Se hace cada vez más necesario desarrollar herramientas y mecanismos que permitan gestionar eficazmente grandes volúmenes de información y racionalizar el acceso de los usuarios de la Web a recursos de su interés. En ámbitos como el académico y científico este problema se hace, si cabe, más complejo y acuciante debido al enorme volumen de literatura que se genera y a las particulares necesidades de información de su comunidad de usuarios (que no siempre se ajustan a los esquemas tradicionales). En las universidades una de las principales vías de acceso a la información son las bibliotecas universitarias, sobre todo aquellas que disponen de una plataforma digital para almacenar y difundir sus recursos. Muchas de ellas, además, ponen a disposición de sus usuarios diversos servicios

---

de valor añadido como servicios de redifusión de contenidos y sistemas de recomendación y filtrado de información (Geisler, McArthur, Giersch 2001) (Huang et al. 2002) que les facilitan la tarea de elegir documentos relevantes que se ajusten a sus necesidades de información y preferencias (muchas veces definidas mediante el uso de perfiles de usuario -(Kuflik, Shoval 2000)-).

No obstante, estos sistemas se limitan a ofrecer recomendaciones desde un enfoque poco flexible en el que se ofrecen al usuario solo recomendaciones sobre documentos que se ajustan fielmente a su perfil, lo cual le permite profundizar en su área de conocimiento, pero obvia otras disciplinas. No es infrecuente, por ejemplo, que este tipo de usuarios desee encontrar recursos relevantes que conecten su área de investigación con otras áreas afines, de forma que les sea posible establecer nuevas líneas de investigación o grupos de trabajo interdisciplinarios. Es evidente que la recomendación que en cada caso debe generar el sistema tendría que ser sensiblemente diferente.

En este trabajo proponemos un sistema capaz de trabajar con ambas políticas de recomendación y que aplica conjuntamente diferentes soluciones tecnológicas que permiten atacar estos problemas de acceso a la información desde diferentes flancos. Por un lado, el sistema se desarrolla sobre una plataforma multi-agente en la que diferentes agentes software actúan de forma autónoma y proactiva, procesando e intercambiando información y resultados con otros agentes (Hendler 2001) (Maes 1994), y asistiendo a los usuarios en tareas de recuperación de información (Brenner, Zarnekov, Witting 1998) (Fazlollahi, Vahidov, Aliev 2000) (Jennings, Sycara, Wooldridge 1998).

No obstante, el principal problema de utilizar agentes es encontrar un protocolo de comunicación lo suficientemente ágil y flexible que permita el intercambio de información entre agentes, y entre usuarios y agentes debido a la gran variedad de formas en que la información puede ser representada en la Web. La aplicación de técnicas de lingüística difusa puede paliar estas barreras de comunicación mediante el uso de etiquetas lingüísticas, resolviendo así el

---

problema de la cuantificación de conceptos cualitativos. Por último, para mejorar la actividad de los sistemas multi-agente utilizamos tecnologías de Web Semántica que permiten describir la información de una manera interpretable tanto por humanos como por los agentes mismos (de forma que estos pueden operar a nivel semántico con los recursos del sistema), y que además disponen de vocabularios para describir sistemas de organización del conocimiento (KOS) que son usados como herramientas para caracterizar el dominio de especialización del sistema y contextualizar la información que circula por él.

Antes de continuar con la descripción del modelo, es necesario aclarar determinados conceptos relativos a los sistemas multi-agente para comprender las bases sobre las que se sustenta el sistema que proponemos en este trabajo.

## 4.2 SISTEMAS MULTI-AGENTE

Los sistemas multi-agente son aquellos en el que cierto número de agentes cooperan e interactúan unos con otros en un entorno distribuido. En la Web, la actividad de un sistema multi-agente consiste en asistir a los usuarios de Internet en sus procesos de acceso a la información por medio de agentes inteligentes distribuidos para encontrar la información que mejor se ajusta a sus necesidades de información (Moukas, Zacharia, Maes 1999) (Lieberman 1999).

Según indica Maes (1995), un agente es un sistema que intenta alcanzar una serie de objetivos predefinidos en un entorno complejo y dinámico. Según sea dicho entorno, podemos establecer una distinción del concepto de agente entre los llamados habitualmente *robots* cuyo entorno es fundamentalmente físico, y los llamados *agentes software* cuyo entorno consiste en computadores y redes. Ambos conceptos comparten una característica fundamental y es que son autónomos, es decir, capaces de operar y decidir por sí mismos la forma de conseguir los objetivos a alcanzar. Sin embargo, como esta característica es de

---



por sí inherente al concepto de agente, un agente autónomo suele denominarse simplemente agente.

Con respecto al término *inteligente*, podemos encontrar diserciones acerca de considerar si un agente es o no inteligente por naturaleza (Petrie 1996). Nosotros los consideraremos como inteligentes, debido a que presentan, en cierto sentido, un comportamiento humano reduciendo así el trabajo a realizar por los usuarios de Internet. Por tanto, los agentes con los que vamos a trabajar, son considerados *agentes inteligentes*. Para poder ser plenamente operativos en la Web Semántica los agentes inteligentes deben reunir una serie de cualidades básicas (Hendler 1999):

- *Deben ser comunicativos*: Debe poseer capacidades para poder comunicarse tanto con los usuarios como con otros agentes en la Web. Esto no supone únicamente comprender la convención de comunicación utilizada, sino que implica que el agente debe tener conocimiento sobre las características del problema a resolver, como las necesidades, objetivos y preferencias del usuario y el contexto en el que se produce un requerimiento determinado. Por esta razón necesitan recurrir a representaciones compartidas de conocimiento (ontologías) para comunicarse de manera eficiente.
  - *Deben ser capaces*: El agente debe tener la capacidad de proporcionar la información que se le requiere y también de ofrecer servicio, es decir, debe tener capacidad para hacer cosas de manera transparente al usuario (como por ejemplo comparar dos productos y ofrecer al usuario el que más se adapta a sus necesidades y requerimientos).
  - *Deben ser autónomos*: El agente debe poder actuar manera proactiva, sin necesidad de que el usuario supervise constantemente los pasos que tiene que dar. Esto implica que el agente debe ser capaz de tomar sus propias decisiones (siempre dentro de los límites impuestos por el usuario).
-

- *Deben ser adaptables*: El agente debe ser capaz de adaptar dinámicamente la información que provee a los requerimientos de los usuarios y a las condiciones específicas del entorno. Esto implica un proceso de análisis y aprendizaje, tanto del entorno como de los propios usuarios, mediante técnicas de *machine learning*.

Como los agentes inteligentes trabajan de forma autónoma y pueden aprender de las acciones de los usuarios, la tecnología de agentes también ha sido aplicada en el ámbito del filtrado de información para reducir la sobrecarga de información en la Web (Maes 1994).

No obstante, los agentes no sólo recuperan y filtran información en el sentido de documentos Web, sino que también gestionan correos electrónicos, listas de noticias, listas FAQ, etc. (Lieberman 1999) (Maes 1995). Debido a que están más cercanos al usuario, son conocidos como *agentes de interfaz* (Maes 1995). Sin embargo, toda la información que estos agentes obtienen y gestionan, proviene de alguna parte. En efecto, hay servidores a través de Internet que proporcionan esos servicios de información, mail, noticias y FAQs. A los agentes más cercanos a estas fuentes de información se les conoce como *agentes de información* (Sycara et al. 1996).

Puesto que los usuarios de Internet pueden acceder a los agentes de interfaz así como a los agentes de información generales, debido a la gran cantidad de dichos agentes, los usuarios se sienten perdidos y sobrecargados de información. Este hecho nos revela la necesidad de una organización entre los agentes, que implica tanto una jerarquía como una arquitectura. Debido a que los elementos que forman parte del proceso de recuperación de información se disponen de forma distribuida, parece sensato considerar una arquitectura distribuida.

Para estos modelos multi-agente distribuidos se han propuesto y analizado diversas arquitecturas, pero para nuestro estudio concreto destacamos la propuesta en (Sycara et al. 1996). En esta arquitectura, además de los antes mencionados agentes de interfaz y de información, los autores consideran un

---

tercer tipo de agentes, los *agentes de tareas*. Estos agentes gestionan los procesos de toma de decisión y el intercambio de información con los agentes de información, resolviendo posibles conflictos que pudieran surgir y realizando procesos de agregación de información, con el objetivo de liberar a los agentes de interfaz de algunas tareas que los hacen menos eficientes. Por lo tanto, este modelo desarrolla la actividad de recuperación contando con cinco niveles: *nivel de usuario, nivel de agente de interfaz, nivel de agente de tareas, nivel de agentes de información y nivel de fuentes de información*.

Sin embargo, la ausencia de conexión y comunicación entre los agentes ha provocado un descenso en la calidad y la conveniencia de la información recuperada, además de en la eficiencia del sistema en las tareas de recuperación y filtrado. Por lo tanto, uno de los aspectos fundamentales a considerar es el diseño de protocolos apropiados de comunicación entre los agentes implicados.

La gran variedad de representaciones y evaluaciones de la información en Internet constituye el principal obstáculo para esta comunicación entre agentes, y el problema es más acusado en los casos en los que los usuarios forman parte del proceso. Este aspecto acentúa la necesidad de una mayor flexibilidad en la comunicación entre los agentes y entre éstos y los usuarios (Yager 1997) (Yager 2001). Para solucionar este problema, se han propuesto varias soluciones como algunos enfoques para gestionar información flexible mediante el uso de información lingüística tanto a nivel de agentes como de usuarios (Yager 1997) (Delgado et al. 2001) (Delgado et al. 2002), y la aplicación de tecnologías de Web Semántica, ya que proporcionan un vocabulario común de intercambio de información que es interpretable tanto por los agentes software como por los usuarios de la Web.

---

### 4.3 MODELO SE SERVICIO SEMÁNTICO-DIFUSO DE DSI PARA BIBLIOTECAS DIGITALES

El modelo que presentamos en esta memoria es fruto de la mejora de un sistema de recuperación de información previo que ha sido modificado para añadir nuevas funcionalidades y servicios. En (Herrera-Viedma et al. 2007) se define un modelo de sistema multi-agente basado en técnicas de modelado lingüístico difuso que mejora la recuperación de información mediante el uso de perfiles de usuario, y aplicando tecnologías de Web Semántica con las que se define una infraestructura semántica sobre la que operan un conjunto de agentes software.

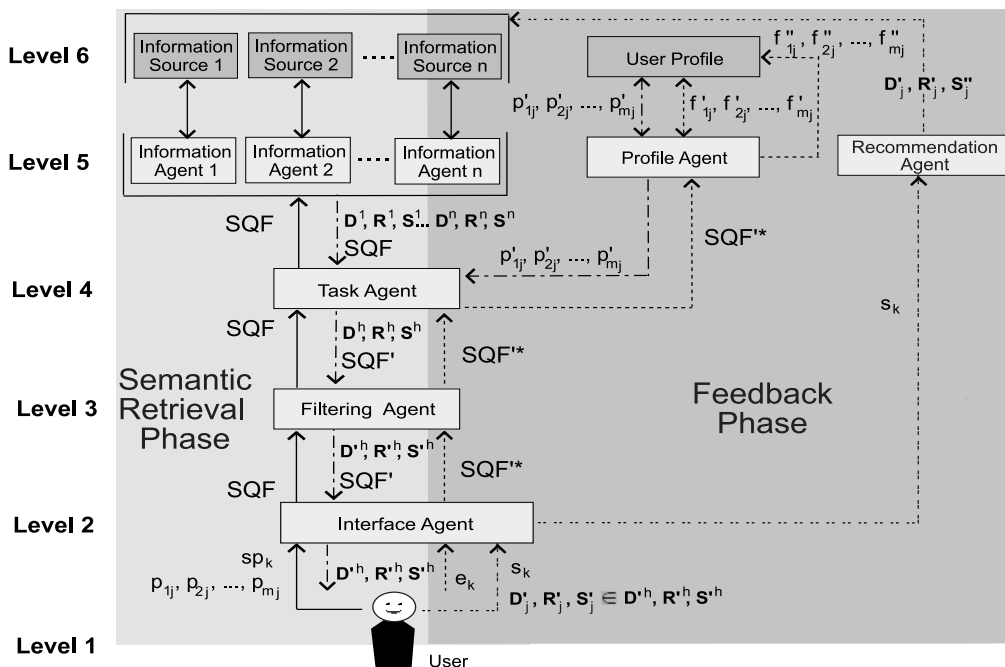


Fig.25 Sistema multi-agente de recuperación y filtrado de información (Herrera-Viedma et al. 2007)

El modelo presenta una estructura jerárquica de seis niveles de actividad (*nivel de usuarios, nivel de agente de interfaz, nivel de agente de filtrado, nivel de agente de tareas, nivel de acceso a la información y nivel de recursos de información*), dos fases de actividad (*fase de recuperación y fase de realimentación*), y un conjunto de agentes (*agente de interfaz, agente de filtrado,*

*agente de tarea, agente de perfiles, agente de recomendación y agentes de información*) que desarrollan sus tareas dependiendo de la fase de actividad.

1. *Fase de recuperación de información:* En esta fase se produce la recuperación y filtrado de recursos y se compone de dos sub-procesos: un primer proceso de búsqueda semántica en el que se realiza una consulta pasiva (de acuerdo a las preferencias definidas en el perfil del usuario) a un repositorio de documentos descritos en formato RDF usando para ello un lenguaje de interrogación semántico, y un segundo proceso de filtrado que consiste en seleccionar los recursos que mejor se adecuan a los requerimientos y restricciones impuestas por los usuarios.
2. *Fase de realimentación o feedback:* Esta fase se corresponde con la actualización de los perfiles de usuario y la generación de recomendaciones colaborativas. En ambos casos es necesaria la intervención activa de los usuarios proporcionando sus valoraciones sobre la eficiencia del sistema para generar recomendaciones y sobre la idoneidad de los recursos recomendados.

Basándonos en la infraestructura básica definida por este modelo presentamos una nueva propuesta que desarrolla un servicio de difusión selectiva de información (DSI) que fundamenta su funcionamiento en la creación de «boletines de novedades» en los que se recoge la descripción básica de los recursos recientemente adquiridos por la biblioteca, o aquellos que por algún motivo puedan ser de interés para sus usuarios. Para definir estos boletines se utiliza RSS 1.0 (Begeed-Dov et al., 2001B), un vocabulario que utiliza la sintaxis y estructura de datos de RDF, y que está diseñado específicamente para gestionar listados de hiperenlaces de una manera sencilla y flexible.

Hemos simplificado la estructura del modelo predecesor definiendo solo 3 agentes (agente de interfaz, agente de tarea y agente de información) que se distribuyen en una arquitectura jerárquica de 5 niveles:

---

- Nivel 1. *Nivel de usuario* donde el individuo se da de alta en el sistema y define un conjunto de preferencias o términos que describen sus necesidades de información que pasarán a formar parte de su perfil de usuario. Estos términos tienen asociado un peso determinado que representa la importancia relativa que para el usuario tiene esa preferencia concreta. El valor de ese peso está representado por una etiqueta lingüística extraída de la variable lingüística “*Nivel de relevancia*”.
  - Nivel 2. *Nivel de interfaz* donde el agente de interfaz (normalmente uno por usuario) actúa como intermediario entre el usuario y el agente de tarea, y realiza diferentes operaciones sencillas de filtrado a petición del usuario.
  - Nivel 3. *Nivel de tarea* donde el agente de tarea (normalmente uno por agente de interfaz) realiza gran parte de las operaciones que se desarrollan en el sistema, como la creación y actualización de perfiles de usuario y canales RSS, o la generación de las alertas bibliográficas.
  - Nivel 4. *Nivel de agentes de información*, donde estos agentes operan directamente con las fuentes de información a las que tiene acceso el sistema, actuando como intermediarios entre las fuentes de información y el agente de tarea.
  - Nivel 5. *Nivel de recursos*, donde se hallan todas las fuentes de información de las que dispone el sistema. Los componentes básicos de este nivel son un repositorio de documentos a texto completo (que conforman el fondo de la biblioteca digital) y una serie de elementos informativos que están descritos utilizando diferentes vocabularios basados en RDF: uno o varios canales RSS que contienen el listado de recursos que se presentan como novedades bibliográficas de la biblioteca; un repositorio de perfiles de usuario donde se almacenan los perfiles que recogen las preferencias y el histórico de
-

recomendaciones de los usuarios; y un tesoro que describe el dominio de especialización de la biblioteca.

La semántica de los diferentes elementos que conforman el modelo (sus características y las relaciones que establecen entre ellos) se define en una o varias ontologías web interoperables desarrolladas en OWL a las que de forma recurrente acudirán los diferentes agentes del sistema para desarrollar sus tareas.

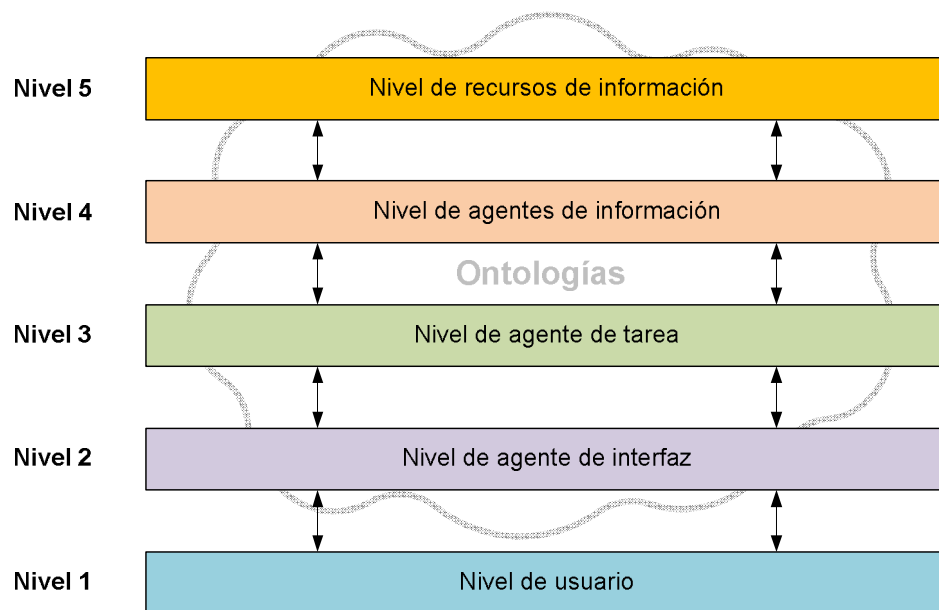


Fig.26 Niveles del modelo de servicio de DSI

El servicio de DSI dispone de 4 módulos funcionales principales mediante los cuales desarrolla su actividad:

- *Módulo de creación de perfiles y canales RSS:* En este módulo los usuarios pueden generar sus perfiles personales proporcionando una serie de datos personales y un conjunto de preferencias (que deben ponderar utilizando etiquetas lingüísticas), y los administradores del sistema definir los canales RSS que van a servir como “boletines de novedades”.

- *Módulo de push de información:* Constituye el servicio de DSI propiamente dicho y en él se gestiona tanto la generación de las alertas bibliográficas, como las recomendaciones colaborativas.
- *Módulo de actualización de perfiles o feedback:* En este módulo, y de acuerdo a las valoraciones del usuario sobre los ítems recomendados por el sistema, se procede a la actualización de su perfil (modificando el peso de las preferencias y el registro histórico de recomendaciones del usuario).
- *Módulo de recomendación colaborativa:* El objetivo de este módulo funcional es generar recomendaciones sobre recursos de acuerdo a las valoraciones previas de expertos afines al usuario activo.

A continuación, describimos de manera pormenorizada cada uno de los elementos principales que conforman el modelo, así como los procesos que se llevan a cabo en los diferentes módulos funcionales.

### **4.3.1 ELEMENTOS DEL MODELO**

Son 4 los elementos básicos que componen el sistema: el tesoro de la biblioteca digital, el repositorio de perfiles de usuario, el repositorio de documentos a texto completo, y los canales RSS. Veamos cada uno de ellos en detalle.

#### **4.3.1.1 TESAURO**

Un elemento imprescindible del servicio de DSI es el tesoro, una herramienta ampliamente utilizada en las bibliotecas tradicionales, que permite describir y estructurar jerárquicamente los conceptos más relevantes del dominio de especialización específico de la biblioteca definiendo las relaciones semánticas que establecen dichos conceptos entre sí (básicamente de equivalencia, jerárquicas y asociativas).

---



La biblioteca al ser universitaria (y por lo tanto, multidisciplinar por naturaleza) deberá disponer de diferentes esquemas de conceptos (tesauros) u ontologías que representen el conocimiento de las distintas áreas que cubre temáticamente. En el caso de que exista más de un área se puede optar por utilizar un único tesoro que englobe todos los dominios de conocimiento que cubre temáticamente la biblioteca, o bien recurrir a una solución más económica que consista en la integración o importación de diferentes tesauros.

Los términos del tesoro conforman el dominio de expresión tanto del campo de clasificación de los ítems del canal RSS, como de las preferencias que representan las necesidades de información de los usuarios.

Para el modelo hemos decidido utilizar un tesoro frente a una ontología ya que, a pesar de que las ontologías permiten definir desarrollos semánticos más profundos para las relaciones entre conceptos, los requerimientos definidos por el sistema admiten el uso de esquemas conceptuales más simples (con una menor capacidad expresiva pero mucho más sencillos de desarrollar y mantener).

#### **4.3.1.2 PERFILES DE USUARIO**

Son representaciones estructuradas que contienen los datos personales, intereses y preferencias de los usuarios, con las que los agentes pueden operar para personalizar el servicio de DSI. En este caso concreto están definidos básicamente con el vocabulario FOAF (Brickley, Miller 2005) (lo cual favorece la interoperabilidad del perfil ya que este es vocabulario que está ampliamente extendido y viene respaldado por una ontología definida en OWL) y un vocabulario propio para definir aquellos campos necesarios para completar la descripción y que no están incluidos en FOAF.

---

Los perfiles se generan en el momento en que el usuario se da de alta en el sistema y se estructuran en 4 partes: un perfil público que incluye datos relativos a su identidad y filiación, un perfil de seguridad donde se definen parámetros que garantizan el acceso seguro a la biblioteca (básicamente un identificador de usuario y una contraseña), un perfil privado donde se recogen sus intereses y preferencias, y un registro histórico con las valoraciones que realizan sobre los recursos recomendados por el sistema.

#### **4.3.1.3 DOCUMENTOS**

El conjunto de documentos disponibles en la biblioteca a texto completo pueden estar almacenados en un repositorio local o bien estar accesibles de forma remota en la red. Dado que el etiquetado semántico de un documento a texto completo es virtualmente imposible con los medios disponibles hoy día, el sistema requiere trabajar con las descripciones que de ellos se almacenan en el canal RSS en forma de ítems. Estas descripciones incluyen básicamente los datos bibliográficos esenciales de cada documento, una serie de materias que describen su contenido y un enlace de acceso al documento a texto completo.

##### **A) TIPOLOGÍA DE LOS DOCUMENTOS**

La definición de tipologías documentales es de gran utilidad a la hora de establecer filtros sobre las consultas al sistema, ya que una vez que el usuario recibe un listado de recursos que concuerdan con las preferencias recogidas en su perfil, podrá añadir niveles adicionales de filtrado definiendo restricciones sobre los tipos de documentos recomendados de acuerdo a sus necesidades.

No obstante, la tipología de documentos que forman parte de una biblioteca está condicionada de una manera determinante por varios factores.

---

<b>Libro:</b> Monografías de carácter científico.	<b>Capítulo de libro:</b> Partes de una monografía de carácter científico.
<b>Bibliografía:</b> Documentos que recogen un listado de recursos seleccionados relativos a un área determinada.	<b>Temario:</b> Documento que recoge el conjunto de temas que forman parte de una asignatura y que puede incluir lecturas relacionadas.
<b>Guía/ Manual:</b> Monografías específicas para instruir de manera amplia sobre un área, herramienta o lenguaje de programación determinado.	<b>Tutorial:</b> Documentos específicos para instruir de forma somera sobre la utilización de una herramienta, lenguaje de programación, etc.
<b>Material de formación de personal:</b> Documentos de carácter técnico y específicos para la formación de personal en una tarea determinada.	<b>Actas de congresos:</b> Documento que recoge todas las ponencias de un congreso.
<b>Ponencia:</b> Comunicación individual presentada en un congreso.	<b>Póster:</b> Tipología específica para los póster presentados a congreso.
<b>Presentación:</b> Documento correspondiente a la presentación de una comunicación a congreso.	<b>Documentos múltiples:</b> Documentos que forman parte de una serie determinada.
<b>Informes:</b> Incluye diferentes documentos como informes técnicos, proyectos y planes estratégicos.	<b>Tesis:</b> Abarca cualquier proyecto que tiene como objetivo la obtención del grado de doctorado o master.
<b>Conjuntos de datos:</b> Principalmente de carácter estadístico y de utilidad para la investigación.	<b>Borradores (Preprints):</b> Documentos que aún no han sido publicados.
<b>Artículo de revista (en línea/ sin paginar):</b> Documentos científicos en formato electrónico, sin paginar y que no tienen equivalente impreso.	<b>Artículo de revista (impreso / paginado):</b> Documento científico en formato electrónico, paginado y que tiene equivalente impreso.
<b>Artículo de prensa:</b> Documentos divulgativos publicados en prensa no especializada.	<b>Revisiones:</b> Comentarios críticos de libros, artículos científicos, etc.
<b>Otros:</b> Cualquier otra tipología de interés no recogida en las definidas en este listado.	

Tabla 4. Tipología documental común en Biblioteconomía y Documentación (Fuente: E-LIS [11])

Uno muy importante es el nivel de conocimientos o competencias de los usuarios. Es evidente que los requerimientos de un estudiante no van a ser las mismas que las de un docente o un investigador, y por lo tanto determinadas tipologías documentales con un alto nivel de especialización (como las tesis doctorales o los informes) no van a ser adecuadas para los usuarios con un menor nivel de conocimientos por lo que deberán recurrir a publicaciones más divulgativas o didácticas).

Por otro lado, la naturaleza de las actividades que realizan las diferentes comunidades de usuarios de una biblioteca también es importante para definir la tipología de documentos que debe conformar el fondo. Así por ejemplo, si nos centramos en las tareas que desarrollan los usuarios de las bibliotecas universitarias (básicamente alumnos, docente e investigadores) podemos definir dos categorías generales de actividades:

- *Tareas de investigación, docentes o de aprendizaje:* Aquellas que de manera específica realizan investigador, docente y estudiante respectivamente para profundizar o ampliar sus conocimientos en un área determinada. La tipología documental que utilicen va a depender directamente de su área de especialización, ya que los canales tradicionales para la comunicación científica pueden diferir de unas áreas de conocimiento a otras (presentándose las diferencias más evidentes entre las áreas científico-técnicas y las de humanidades) (Carrizo Sainero 1994). Por poner un ejemplo, en el área de la Biblioteconomía y la Documentación una posible selección de tipos documentales adecuada para estos profesionales podría ser la que se recoge en E-LIS [11] (Tabla 4). No obstante, esta no tiene por qué ser necesariamente la selección la más adecuada en otras áreas de conocimiento.
-

- *Tareas de financiación:* Son aquellas que van encaminadas a la búsqueda de fuentes que soporten económicamente proyectos que ayuden a desarrollar o complementar la actividad investigadora, docente o de formación. En este caso los recursos necesarios serán las convocatorias de becas y contratos con cargo a grupos o proyectos, y otras convocatorias de carácter general como ayudas, proyectos, becas, contratos y premios.

El sistema debe ser, por lo tanto, capaz de distinguir entre diferentes tipologías documentales para ofrecer al usuario la capacidad de obtener niveles adicionales de filtrado.

#### 4.3.1.4 CANALES RSS

En los últimos años, gracias a la popularización de las bitácoras, también conocidas como *weblogs* o *blogs*, se está imponiendo el uso de varios vocabularios específicos para la redifusión o *sindicación* de contenidos (es decir, hacer accesible a otros usuarios de Internet el contenido de un sitio web mediante listas de enlaces denominadas canales o *feeds*). Para crear nuestros boletines de novedades utilizamos RSS 1.0 (*RDF Site Summary*), un vocabulario que se basa en el modelo de datos y sintaxis de RDF/XML (lo cual hace que su estructura de datos sea más compleja). No obstante, aunque las versiones 0.9x y 2.0 (*Really Simple Syndication*) utilizan sintaxis XML para definir un vocabulario simple y fácil de manejar (esto ha permitido que su uso esté ampliamente extendido en la red), RSS 1.0 dispone de mayores capacidades de interoperabilidad y extensibilidad gracias al uso de módulos.

Estos módulos definen etiquetas específicas que permiten ampliar el vocabulario sin necesidad de modificar el núcleo de la especificación cada vez que se quieran añadir nuevos elementos descriptivos. En este modelo se utilizan el módulo *DC* (Begeed-Dov et al. 2000) para definir la información bibliográfica básica de los ítems utilizando los elementos establecidos por la DCMI [130], y el módulo *syndication* (Begeed-Dov et al. 2001A) para facilitar a los agentes la sincronización y la actualización del canal. A continuación, analizamos cada uno de estos módulos de una manera más exhaustiva.

##### A) MÓDULO DUBLIN CORE

El módulo *Dublin Core* define quince elementos que se corresponden con los quince elementos metadata fundamentales del vocabulario *Dublin Core* que pueden ser utilizados para especificar la información contenida en los diferentes elementos del canal RSS (canal, imagen, ítems y entrada de texto).

---

Estos elementos son:

- *<dc:title>*: Nombre que identifica al recurso.
  - *<dc:creator>*: Responsable principal de la autoría del recurso.
  - *<dc:subject>*: Define la materia del recurso.
  - *<dc:description>*: Resumen o reseña del contenido.
  - *<dc:publisher>*: Responsable de la edición o publicación del recurso.
  - *<dc:contributor>*: Responsable (no principal) que contribuye a la autoría del recurso.
  - *<dc:date>*: Periodo temporal asociado al recurso. Se representa utilizando el formato estandarizado W3CDTF [170] que sigue el patrón de formato YYYY-MM-DDThh:mm:ss.sTZD, donde Y son dígitos para definir el año, M dígitos para el mes, D dígitos para el día, T es un carácter que introduce el formato de hora donde h es un dígito para las horas, m para los minutos y s para los segundos (incluyendo fracciones de segundo). TZD es un identificador de zona horaria cuyos valores pueden ser Z (cuando se define la hora universal coordinada), o las horas y minutos de adelanto (-hh:mm) o atraso (+hh:mm) con respecto a la hora universal coordinada. Un ejemplo válido para este campo sería 2007-12-20T10:28:45.23+01:00.
  - *<dc:type>*: Género o naturaleza del recurso.
  - *<dc:format>*: Formato, soporte o dimensiones del recurso.
  - *<dc:identifier>*: Referencia que permite identificar unívocamente un recurso.
  - *<dc:source>*: Recurso del que deriva el recurso que estamos describiendo.
  - *<dc:language>*: Idioma en que viene expresado el recurso.
-

- *<dc:relation>*: Especificación de un recurso asociado al que estamos describiendo.
- *<dc:coverage>*: Definición del ámbito o contexto temporal y geográfico (o espacial) del recurso.
- *<dc:rights>*: Información sobre los derechos e autor a los que está sujeto el recurso a describir.

En todos los casos, el uso de los elementos *Dublin Core* es el estándar, añadiendo un valor literal a las diferentes propiedades definidas. No obstante, en el caso del elemento *dc:subject* se está discutiendo la posibilidad de enriquecer su semántica añadiendo el módulo taxonómico *taxonomy* de forma que las materias estén definidas como la descripción de elementos de una taxonomía cuyo valor literal viene especificado por un constructor *rdf:value*.

#### B) MÓDULO SYNDICATION

Por su parte el módulo *syndication* extiende el vocabulario con tres propiedades que se pueden añadir al elemento *<channel>* especificar su frecuencia de actualización y de esta forma facilitar a los agentes la programación de la búsqueda de nuevos ítems en el canal.

- *<updatePeriod>*: Define la unidad temporal usada para cuantificar el lapso de de tiempo entre actualizaciones. Sus valores admitidos son *yearly*, *monthly*, *weekly*, *diary*, y *hourly* (para actualizaciones anuales, mensuales, semanales, diarias u horarias, respectivamente).
  - *<updateFrequency>*: Cuantifica el número de unidades temporales entre actualizaciones. Su valor es un entero positivo, y por defecto es 1.
  - *<updateBase>*: Esta propiedad permite definir una fecha y hora que se toman como referencia para poder operar con las propiedades
-



<updatePeriod> y <updateFrequency>. Su valor toma el formato de fecha W3CDTF.

La estructura de los canales comprende dos áreas: una en la que se describe el canal en sí y otra en la que se listan los diferentes recursos o ítems. La descripción del canal necesita una serie de datos básicos como su título, una breve descripción del contenido o su frecuencia de actualización. Por su parte, las descripciones de los ítems necesitan los campos título, autor, resumen, enlace al recurso, fecha de creación del ítem y clasificación por materias.

```
<item rdf:about="http://.../ies/Bib">
  <title>Biblioteconomía aplicada</title>
  <link>http://.../ises/Bib.pdf</link>
  <description>Resumen del recurso</description>
  <dc:creator>Valeria Allesi</dc:creator>
  <dc:date>2005-03-17T13:25:42Z</dc:date>
  <dc:subject>Biblioteconomía</dc:subject>
</item>
```

Fig.27 Ejemplo básico de ítem del canal RSS

---

### 4.3.2 MÓDULOS DEL SISTEMA

A continuación describimos los principales módulos funcionales que forman el sistema y los procesos que se desarrollan en ellos. Módulo de creación de perfiles de usuario y canales RSS, Básicamente son cuatro: el de creación de perfiles y canales RSS, el de actualización de perfiles, y los de generación de alertas y recomendaciones cooperativas. Pasemos a continuación a describirlos pormenorizadamente.

#### 4.3.2.1 MÓDULO DE CREACIÓN DE PERFILES Y CANALES RSS

##### *a) Creación de canales*

Los usuarios al darse de alta en la biblioteca cumplimentan un formulario a través del cual el sistema obtiene una serie de datos personales de identificación básicos que pasarán a formar parte del perfil público, y un *login* y una contraseña que serán almacenados en su perfil de seguridad. Seguidamente, el individuo debe especificar una serie de palabras clave o conceptos expresados en lenguaje natural que a su juicio definen de una manera adecuada sus necesidades de información.

El sistema procede a continuación a equiparar léxicamente estos conceptos con los términos del tesoro de la biblioteca utilizando como medida de similitud el algoritmo *edit tree* o distancia Levenshtein (Levenshtein 1966). Esta función equipara cadenas de caracteres y devuelve el mismo término introducido cuando la equiparación es exacta, o el término léxicamente más similar en caso contrario.

En caso de no existir una equiparación positiva entre, el sistema sugiere un nuevo conjunto de descriptores. Si los términos sugeridos son satisfactorios para el usuario serán directamente añadidos al perfil (una vez el usuario los haya ponderado convenientemente). En caso de que los términos no sean satisfactorios, el sistema dispondrá de mecanismos para que el usuario pueda

---

navegar por el tesoro y de esta forma seleccionar por él mismo los términos que a su parecer son más adecuados. Un ejemplo de este tipo de aplicaciones es ThManager [171], un proyecto de la Universidad de Zaragoza que permite editar, visualizar y recorrer estructuras definidas en SKOS Core.

Como acabamos de comentar, cada uno de los términos seleccionados por el usuario para definir sus áreas de interés tienen asociado un peso determinado (etiquetado como *<relev>*), que representa el grado de interés que suscita en el usuario esa materia concreta.

Estos pesos son un elemento clave en el funcionamiento del sistema, debido a varias razones:

- son un factor determinante a la hora de calcular tanto la relevancia de los ítems del canal RSS con respecto a las preferencias de los perfiles de los usuarios, como la semejanza entre perfiles de usuario.
- el proceso de actualización de perfiles se centra en recalcular estos pesos.

El rango de posibles valores para estos pesos está definido por un conjunto de 7 etiquetas lingüísticas que se extraen de la variable lingüística «*Nivel de relevancia*» y cuyo dominio de expresión es  $S = \{nulo, muy\ bajo, bajo, medio, alto, muy\ alto, total\}$ .

El área del perfil correspondiente al registro histórico de recomendaciones no se genera en esta fase, sino en la fase de actualización de perfiles, por lo que será analizada más adelante.

---

### *b) Creación de canales RSS*

Para el proceso de creación, el administrador dispone de un interfaz a través del cual genera la descripción del canal en sí como su título, un resumen del contenido del canal y la frecuencia con la que los agentes deben actualizarlo. A este almacén se añaden posteriormente las descripciones de los ítems durante el proceso de actualización. Periódicamente, el agente de tarea comprueba el repositorio de documentos en busca de aquellos que se han almacenado recientemente para generar su correspondiente representación en el canal RSS.

Si existe en la Web una fuente de la que se pueda extraer la información necesaria para completar la descripción de estos ítems (una base de datos o repositorio de acceso público) serán los agentes de información los encargados de obtenerla para que el agente de tarea genere la descripción para cada registro (consignando título, autor, resumen y enlace al recurso). En el caso de que no se pueda extraer de la Web esta información de una manera sistemática y automatizada deberán ser los administradores del sistema los responsables de generar la descripción completa de los ítems.

No obstante, en ambos casos será necesario que sea precisamente el administrador quien defina las materias que describen el contenido del recurso. Para facilitar esta tarea dispondrán de una herramienta que les asiste en el proceso de asignar materias a los ítems que funciona de forma análoga a cómo seleccionan los usuarios las preferencias que forman parte de su perfil: el administrador propone una serie de materias que son equiparadas léxicamente con los términos del tesoro utilizando el algoritmo *edit tree* (Levenshtein 1966).

Aquellos casos en los que la equiparación sea exacta los términos serán asignados como materia (junto a sus sinónimos), y en caso contrario el sistema sugerirá una serie de términos léxicamente similares que el administrador puede utilizar o no, siguiendo su propio criterio.

---

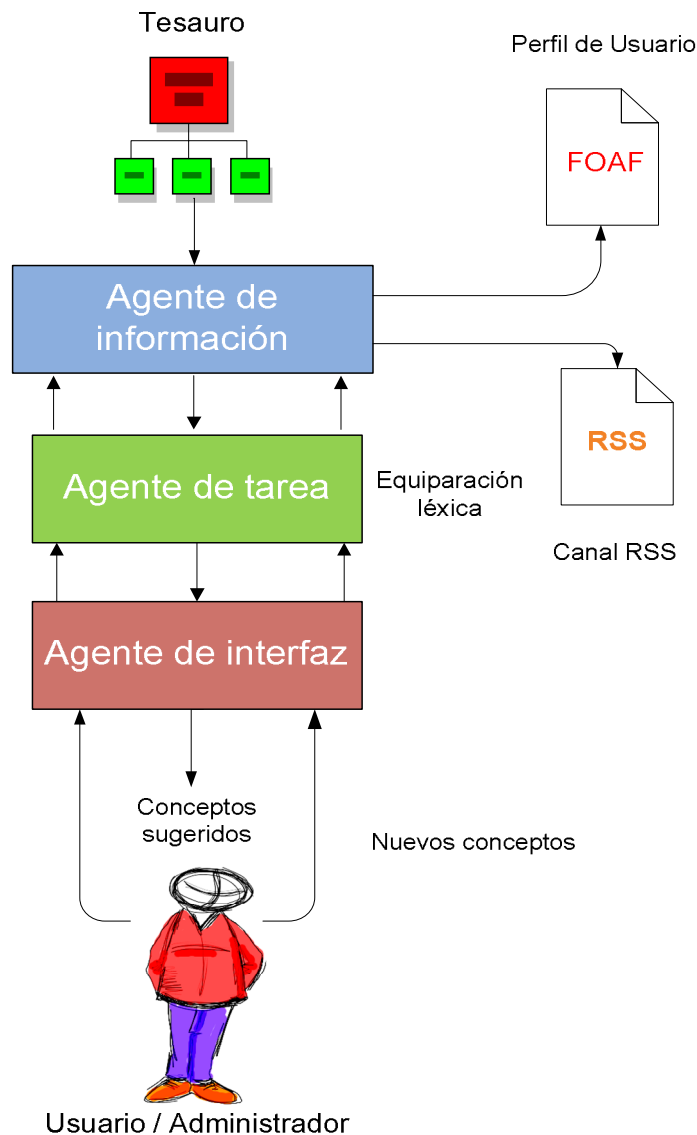


Fig.28 Proceso de creación de perfiles y canales RSS

#### 4.3.2.2 MÓDULO DE GENERACIÓN DE ALERTAS O *PUSH* DE INFORMACIÓN

En este proceso es en el que realiza la función de DSI propiamente dicha. Consiste básicamente en realizar, en lugar del usuario, una consulta pasiva al canal RSS de la biblioteca sobre las áreas de interés del usuario definidas en su perfil, y posteriormente generar una alerta personalizada que es presentada al usuario sin que este lo haya requerido expresamente (lo que se conoce como

*push* o recepción pasiva de información). El proceso se desarrolla en los siguientes pasos:

*Paso 1:* El usuario accede al sistema de manera identificada introduciendo su nombre de usuario y clave de acceso.

*Paso 2:* A continuación, el agente de tarea procede a equiparar las áreas de interés contenidas en el perfil del usuario activo con los descriptores que caracterizan el contenido de los  $n$  ítems del canal RSS y de esta forma discriminar aquellos que mejor se adaptan a las necesidades de información concretas de ese individuo. En este caso, en lugar de utilizar la tradicional equiparación léxica, en la que se comparan las cadenas de caracteres de dos términos, vamos a utilizar una medida de semejanza semántica aprovechando el tesoro del sistema como herramienta de organización del conocimiento. Para ello utilizamos la función que permite calcular la semejanza de objetos RDF presentada en (Oldakowsky, Byzer 2005). Esta función permite determinar la distancia entre dos conceptos de una taxonomía de acuerdo a su situación dentro de la jerarquía de conceptos.

La semejanza se define como:

$$sim_c(c_1, c_2) = 1 - d_c(c_1, c_2)$$

La distancia entre dos conceptos dados representa el camino de uno a otro a través del antecesor común más cercano (*closest common parent o ccp*) y se calcula así:

$$d_c(c_1, c_2) = d_c(c_1, ccp) + d_c(c_2, ccp)$$

$$d_c(c, ccp) = milestone(ccp) - milestone(c)$$

donde a cada concepto de la taxonomía se le asigna un puntero o hito (*milestone*). Este puntero se puede calcular utilizando tanto una función lineal

como una exponencial dependiendo de las características y necesidades de nuestro sistema. Si optamos por definir el puntero lineal, este se calculará según la siguiente fórmula:

$$milestone(n) = 1 - [l(n) / l(N)]$$

donde  $l(n)$  es la profundidad del nodo  $n$  en la jerarquía y  $l(N)$  representa el nivel jerárquico más profundo. Si nos decidimos por utilizar el puntero exponencial se aplicará la función definida en (Zhong et al. 2002):

$$milestone(n) = 1 / 2k^{l(n)}$$

donde  $k$  es un factor mayor que uno que indica la proporción en que el valor del hito decrece conforme bajamos niveles en la jerarquía. Su valor puede variar dependiendo de la profundidad de la taxonomía.

*Paso 3:* Una vez definida la similaridad entre preferencias y materias es posible calcular la relevancia de un recurso determinado con respecto al perfil de un usuario específico. No obstante, para ello es necesario recurrir al concepto de “solapamiento semántico”.

El concepto de *solapamiento semántico* intenta solventar el problema del cálculo de la similaridad utilizando operadores taxonómicos ya que a la hora de calcular la similaridad entre conceptos de un mismo tesoro no va ser posible encontrar términos cuya similaridad sea cero, lo cual provoca la aparición de valores de relevancia mayores que 1 y que difícilmente pueden ser normalizados

La idea que subyace a este concepto es la determinación de áreas de máxima intersección entre conceptos. Veámoslo con un ejemplo.

---

Supongamos que queremos comparar el perfil  $P_1$  del usuario activo con un perfil  $P_2$ , donde:

$$P_1 = [p_1, p_2, \dots, p_N] \text{ y } P_2 = [p'_1, p'_2, \dots, p'_M].$$

Gráficamente podríamos representar estas preferencias como curvas suaves cuya superficie es de 1 unidad de área que se solapan de acuerdo al grado de similitud semántica que existe entre ellas.

En este caso concreto supongamos que  $N=3$ ,  $M=2$  y que las preferencias se corresponden semánticamente de la siguiente forma:  $p_1 = p'_1$  y  $p_2 = p'_2$ . El gráfico que obtendríamos sería similar a este (para una mayor claridad hemos supuesto que las preferencias que pertenecen a un mismo perfil no son similares entre sí):

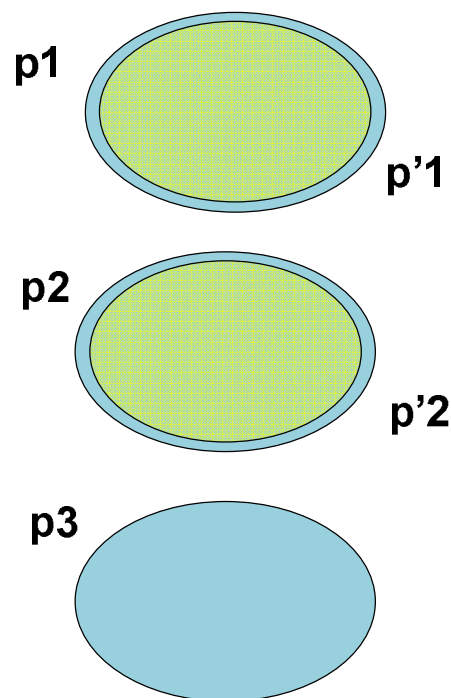


Fig.29 Solapamiento semántico (I)

En este caso, hay 2 preferencias iguales entre sí y por lo tanto se definen 2 áreas de intersección donde el nivel de solapamiento es total. En este caso concreto, de las 6 medidas de similitud entre preferencias fruto del producto cartesiano entre las preferencias de ambos perfiles, solo dos de ellas



( $Sim(p_1, p'_1)$  y  $Sim(p_2, p'_2)$ ) definen un solapamiento semántico, que en este caso es completo (es decir, uno de los perfiles está completamente recogido en el otro).

Para calcular la similaridad entre ambos definimos la siguiente función:

$$Sim(P_i, P_j) = \sum_{k=1}^{MIN(N,M)} \frac{H_k(Sim p_i, p_j)}{MAX(N, M)}$$

Donde  $H_k(Sim p_i, p_j)$  es una función que obtiene las  $k$  similaridades máximas definidas entre las preferencias de  $P_i$  y  $P_j$ . En este caso el valor de  $Sim(P_1, P_2) = 1$ .

No obstante, esta situación es una situación ideal ya que lo que ocurre comúnmente es que entre las preferencias existe solo un cierto grado de similitud o solapamiento (si es que este existe). Un gráfico más realista sería uno como el siguiente:

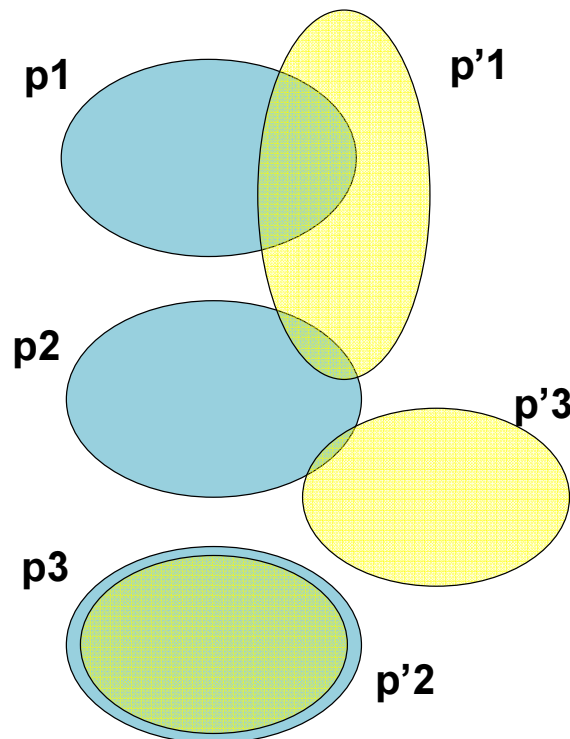


Fig.30 Solapamiento semántico (II)

Aquí, las preferencias que tienen algún tipo de relación semántica entre sí aparecen en el gráfico solapadas proporcionalmente a su grado de relación.

En este ejemplo  $N=M=3$  y de acuerdo a la definición de la función de relevancia, se deberían obtener 3 áreas de solapamiento máximo. Suponiendo los siguientes valores de similaridad:

$$\begin{aligned} Sim(p_1, p'_1) &= 0.30 & Sim(p_2, p'_3) &= 0.05 \\ Sim(p_2, p'_1) &= 0.17 & Sim(p_3, p'_2) &= 1 \end{aligned}$$

la similaridad en este caso sería  $Sim(P_1, P_2) = 0.49$

En el caso de que comparar las preferencias de dos perfiles, además debemos tener en cuenta el peso de ponderación asociado a cada una de ellas. Bastaría con multiplicar las respectivas similaridades por la media de los pesos  $\omega_i$  y  $\omega_j$  asociados a cada par de preferencias.

$$Sim(P_i, R_j) = \frac{\sum_{k=1}^{MIN(N,M)} H_k(Sim p_i, r_j) \left( \frac{\omega_i + \omega_j}{2} \right)}{MAX(N, M)}$$

En el caso de la equiparación entre perfiles  $P_i = \{p_1, \dots, p_N\}$  y recursos  $R_j = \{r_1, \dots, r_M\}$ , las similaridades estarían multiplicadas por el peso asociado a la preferencia, ya que las materias de los ítems no están ponderadas.

$$Sim(P_i, R_j) = \frac{\sum_{k=1}^{MIN(N,M)} H_k(Sim p_i, r_j) \omega_i}{MAX(N, M)}$$

El uso del solapamiento semántico para realizar medidas de relevancia entre términos pertenecientes a un mismo esquema supone asumir una pérdida de información ya que se desechan similitudes entre pares de elementos que no siempre tienen por qué ser residuales. No obstante, se facilita el cálculo de la relevancia y se gana en coherencia al restringir el rango de posibles valores al intervalo  $[0, 1]$ .

*Paso 4:* Una vez determinada la similaridad y la relevancia de los documentos con respecto al perfil, el agente de interfaz presentará al usuario los recursos cuyo valor de relevancia iguale o sobrepase un umbral predefinido  $k$  (con un valor cercano a 1), desechándose cualquier documento que no lo alcance.

*Paso 5:* En el último paso, el agente de interfaz genera una alerta en la página de bienvenida del sitio Web, que advierte al usuario de la existencia de nuevos documentos que pueden resultar de su interés. Esta alerta enlaza directamente con el listado de recursos, desde el cual el usuario puede realizar tareas de filtrado imponiendo restricciones sobre la tipología documental de los recursos y acceder a ellos a texto completo. En el caso de existir más de un canal en la biblioteca, los ítems recuperados de cada canal son agregados por el agente de tarea en un único listado. Si no se hallasen ítems que satisfagan la consulta se notifica al usuario que no hay novedades.

---

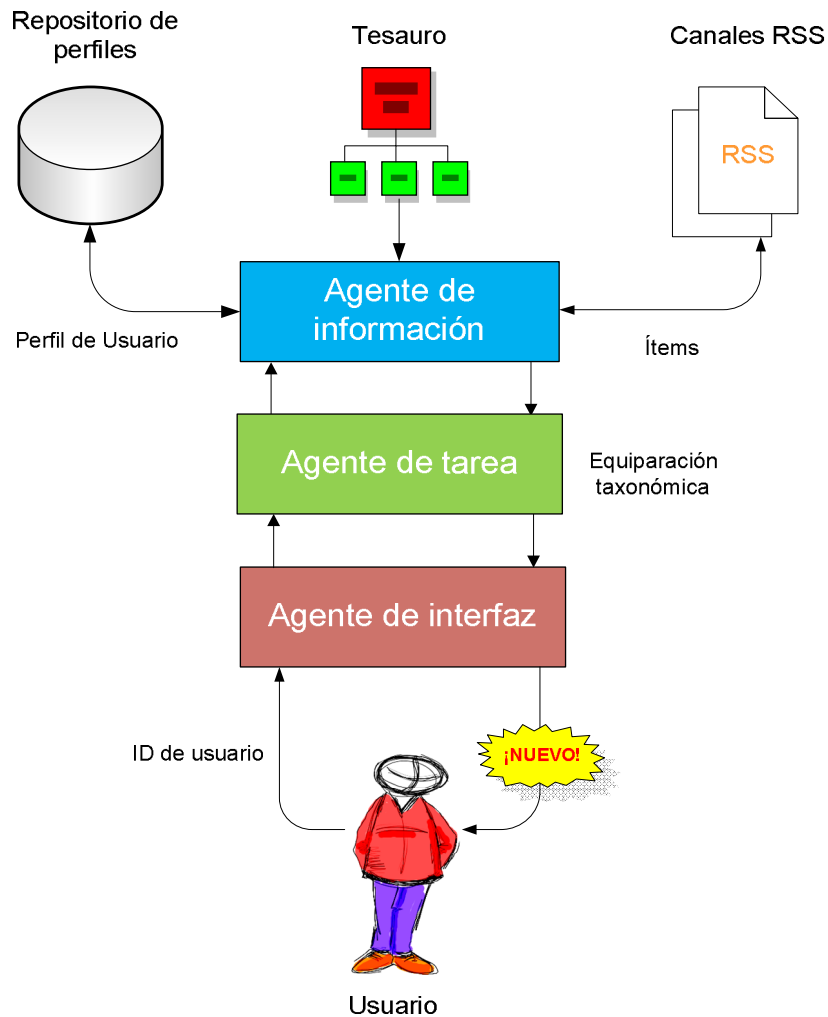


Fig.31 Proceso de generación de alertas

#### 4.3.2.3 MÓDULO DE ACTUALIZACIÓN DE PERFILES O FEEDBACK

Dado que el servicio de DSI se basa en generar consultas pasivas a los canales RSS a partir de las preferencias almacenadas en el perfil de los usuarios, la actualización de estos es fundamental ya que aunque en los perfiles se recogen las preferencias a largo plazo de los individuos, no cabe duda de que estas sufren modificaciones en el tiempo que el sistema debe ser capaz de detectar para ofrecer en cada momento un servicio lo más ajustado posible a las necesidades reales del usuario.

En nuestro modelo la actualización de los perfiles se gestiona mediante un sencillo mecanismo que utiliza técnicas de lingüística difusa y se basa en el *feedback* que proporciona el usuario al evaluar los ítems que le han sido recomendados por el sistema.

Asumiendo las premisas definidas en la teoría del modelado lingüístico ordinal, podemos construir el mecanismo de actualización de perfiles. Este se basa en la actualización del peso de importancia asociado a cada una de los términos de preferencia definidos por el usuario de acuerdo al nivel de satisfacción  $e_j$  que se extrae de la variable lingüística «*Nivel de satisfacción*» y cuyo dominio de definición es  $S' = \{nulo, muy\ bajo, bajo, medio, alto, muy\ alto, total\}$ .

Para ello es necesario definir una función de equiparación similar a las empleadas para modelar pesos de umbral en consultas ponderadas (Herrera-Viedma 2001) que bonifica los pesos de los valores de preferencia presentes en recursos valorados como satisfactorios y las penaliza cuando esta valoración es negativa. Sea  $e_j \in S'$  este grado de satisfacción, y  $\omega_{li}^j \in S$  la frecuencia de la propiedad  $i$  (en este caso  $i$  = «Preferencia») con valor  $l$ , entonces definimos la función de actualización  $g: S' \times S \rightarrow S$ :

$$g(e_j, \omega_{li}^j) = \begin{cases} S_{Min}(a + \beta, T) & \text{si } s_a \leq s_b \\ S_{Max}(0, a - \beta) & \text{si } s_a > s_b \end{cases}$$

$$s_a, s_b \in S \mid a, b \in H = \{0, \dots, T\}$$

donde, (i)  $s_a = \omega_{li}^j$ ; (ii)  $s_b = e_j$ ; (iii)  $a$  y  $b$  son los índices de las etiquetas lingüísticas cuyo valor oscila entre 0 y  $T$  (siendo  $T$  el número total de etiquetas de  $S$  menos uno), y (iv)  $\beta$  es un valor de bonificación que recompensa o penaliza las frecuencias y que se define así:

$$\beta = \text{round} \left( \frac{2|b-a|}{T} \right)$$

donde *round* es la función de redondeo típica.

No obstante, la actualización de los pesos solo se realizará sobre aquella preferencia del perfil del usuario cuya similaridad (no ponderada) con las materias del ítem del canal RSS haya sido la más alta. De esta manera, pretendemos capturar y descubrir las preferencias implícitas del usuario.

Este mecanismo se vuelve más preciso cuanto más recursos evalúe el usuario ya que de esta forma será más fácil para el servicio de DSI «aprender» a seleccionar qué recursos le van a resultar interesantes al usuario de acuerdo a sus preferencias en un momento determinado. Este proceso de evaluación realizado por el usuario no solo sirve para actualizar los perfiles de usuario ya que, como vamos a ver en el siguiente proceso, el *feedback* proporcionado por el individuo puede ser reutilizado para crear un sistema de recomendaciones en el que se saque provecho de su experiencia y capacidad crítica para beneficiar a la colectividad de usuarios de la biblioteca.

En el área correspondiente al registro histórico de recomendaciones se almacenan las valoraciones que el usuario realiza sobre los documentos recomendados por el sistema. Con estas valoraciones el sistema es capaz de generar recomendaciones colaborativas basadas en la opinión de expertos con perfiles similares.

Cada una de las entradas del registro histórico está compuesta por el valor de satisfacción  $e_j$  que el usuario define para ese recurso, la URI que identifica a ese usuario, y una fecha de registro.

El dominio de expresión de las recomendaciones está definido por el siguiente conjunto de 7 etiquetas que se extraen de la variable lingüística «Nivel de satisfacción»:  $S' = \{nulo, muy\ bajo, bajo, medio, alto, muy\ alto, total\}$ .

---

#### 4.3.2.2 MÓDULO DE RECOMENDACIÓN COLABORATIVA

El sistema puede además ofrecer información adicional sobre los recursos recomendados independientemente de la similitud de estos con el perfil del usuario. Por ello, junto al sistema de recomendaciones basado en contenido, el modelo dispone de un sistema de recomendaciones cooperativo basado en las recomendaciones emitidas por otros usuarios de la biblioteca con un perfil similar o afín al del usuario activo. El proceso se desarrolla de la siguiente manera:

*Paso 1:* El sistema realiza un proceso de *clustering* o clasificación de perfiles de usuario para hallar aquellos individuos cuyas preferencias son más similares a las del usuario activo. La semejanza entre perfiles se calcula tal y como se ha explicado en el apartado 4.3.2.2.

*Paso 2:* Una vez definido el conjunto de usuarios similares al usuario activo el agente de tarea procede a comprobar si en el área de registro histórico de valoraciones de sus perfiles respectivos existen entradas en las que se valore alguno de los recursos que van a ser recomendados. En caso afirmativo se procede a agregar las valoraciones lingüísticas utilizando el operador lingüístico difuso LOWA (Herrera, Herrera-Viedma 1997) cuya salida es una nueva etiqueta lingüística que se extrae de la variable lingüística «*Nivel de satisfacción*» y cuyo dominio de definición es  $R' = \{nula, muy\ baja, baja, media\ alta, muy\ alta, total\}$ .

Adicionalmente, y disponiendo del nivel de conocimientos de cada usuario (o su tipología) también sería posible añadir un nuevo nivel de filtrado al sistema que consistiría en generar la recomendación colaborativa a partir de las opiniones de usuarios afines (con intereses similares) pero de la misma tipología o con un nivel de conocimientos similar.

---

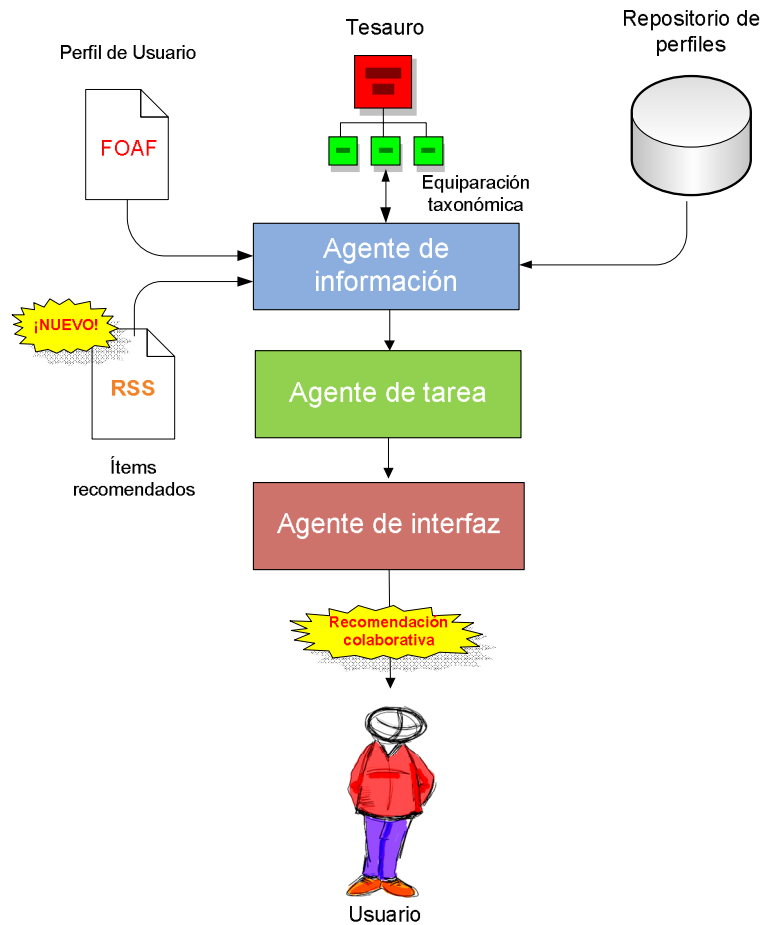


Fig.32 Proceso de recomendación colaborativa

De esta manera conseguimos que cada usuario obtenga recomendaciones adecuadas a sus intereses, pero también a su nivel de formación (evitando la frustración que supone para el usuario acceder a recursos que aparecen como muy recomendables pero cuya complejidad es inadecuada para su nivel formación).

Todas estas cuestiones se han tenido en consideración y han sido aplicadas en el diseño del prototipo que se presenta a continuación, que permite demostrar la viabilidad de aplicación de las propuestas de partida.





---

## CAPÍTULO 5

# D-FUSSION: PROTOTIPO DE SERVICIO SEMÁNTICO-DIFUSO DE DSI PARA BIBLIOTECAS DIGITALES UNIVERSITARIAS

En este capítulo describimos el prototipo de servicio semántico-difuso de servicio de DSI para bibliotecas digitales universitarias D-FUSSION que hemos desarrollado para comprobar la viabilidad del modelo que hemos presentado en el capítulo anterior.

### 5.1 INTRODUCCIÓN

D-Fussion es un prototipo, desarrollado utilizando tecnologías de Web Semántica y técnicas de modelado lingüístico, que tiene como objetivo generar alertas personalizadas sobre recursos de la biblioteca digital que puedan resultar de interés para los usuarios de la misma de acuerdo al perfil que sobre ellos mantiene el sistema.

A continuación pasamos a describir la filosofía de recuperación de información que subyace al desarrollo del prototipo, los elementos componentes, los módulos del modelo implementados, un ejemplo operacional y la arquitectura de la aplicación.

#### 5.1.1 NUEVOS ENFOQUES DE ACCESO A LA INFORMACIÓN

Como ya hemos comentado con anterioridad en este trabajo, el consumidor de información actual en la Web (ya sea un mero usuario de internet o el usuario de una biblioteca digital universitaria) es cada vez más exigente y requiere de

---

medios de acceso a la información cada vez más precisos. Esto ha provocado que en los últimos años se hayan multiplicado las aproximaciones teórico-prácticas al proceso de acceso a la información en un intento de superar el modelo de los buscadores generalistas tradicionales y de esta manera ofrecer soluciones personalizadas prácticamente para cada tipo de usuario o para cada problema concreto que se desea resolver. Así, por ejemplo, Ng (2008) distingue hasta seis maneras distintas de buscar, descubrir y recuperar información en la Web:

1. *Búsqueda basada en un trabajo intelectual apriorístico (human-powered search)*: Esta búsqueda es la que se realiza sobre directorios como Yahoo! o en enciclopedias colaborativas como Wikipedia, en los que la navegación y la forma en que se interrelacionan los contenidos se basa en el trabajo intelectual desarrollado por un equipo humano. Estos directorios suelen tener una pequeña cobertura temática, pero representan una solución complementaria a los buscadores horizontales tradicionales (generalistas) por el alto valor y calidad de sus contenidos.
  2. *Búsqueda centrada en la personalización*: Consiste en adaptar los resultados que se devuelven al usuario de acuerdo a los intereses y preferencias que aparecen recogidos en su perfil personal. Este enfoque definido en un contexto independiente de dominio presenta como inconveniente que el historial de consultas de un usuario, por lo general, no es de gran ayuda para predecir futuras consultas ya que la naturaleza de estas es muy variable. No obstante, en determinados ámbitos verticales (especializados en un nicho o dominio específico) sí que es posible generar mecanismos que capturen el contexto y puedan predecir el comportamiento del usuario con mayor exactitud.
  3. *Búsqueda social*: La idea de este enfoque consiste en refinar las consultas propias de un usuario determinado con el historial de búsqueda de un amigo o usuario afín. Como inconveniente encontramos que si los historiales de dos usuarios afines representan necesidades de
-

información poco estables en el tiempo las ventajas de usar esta perspectiva se reducen considerablemente. No obstante, por lo general este tipo de enfoque tiene como objetivo más el descubrimiento de recursos que definir búsquedas puras, por lo que se podría considerar como un enfoque complementario al de la búsqueda tradicional.

4. *Búsqueda en motores verticales*: Los motores verticales ofrecen buenos resultados para dominios muy concretos, pero eso implica prácticamente la existencia de un motor de búsqueda específico para resolver los problemas de acceso a la información en un área de conocimiento determinada. Esto, no obstante, no representa necesariamente una desventaja ya que utilizar este tipo de motores supone usar una herramienta específica para resolver determinados problemas bien definidos y contextualizados, y no herramientas *fit-to-all* (como la mayoría de las disponibles hoy día) que pretenden dar solución a problemas de acceso a la información de muy diversa naturaleza.
  5. *Procesamiento del lenguaje natural*: Presenta muchos problemas sobre todo cuando las consultas están mal formuladas (de manera imprecisa o vaga) o cuando los documentos no están correctamente redactados. Sin embargo, de nuevo, en dominios controlados este tipo de técnicas permiten obtener unos resultados de precisión y eficiencia mucho mayores que en un ámbito no especializados.
  6. *Búsqueda semántica*: Como hemos visto en esta memoria, consiste en la recuperación de información mediante el uso de tecnologías de Web Semántica y puede ser un complemento para todas las perspectivas de búsqueda aprovechando el potencial de RDF y las ontologías web. De nuevo en un dominio vertical es donde esta técnica tendría mayores posibilidades de mostrar todo su potencial. El inconveniente que presenta este tipo de enfoque es el gran esfuerzo y enorme cantidad de recursos económicos, de tiempo y de personal que conlleva su desarrollo.
-

Esto lleva a muchos desarrolladores a no estar dispuestos a compartir de forma altruista ese trabajo con el resto de usuarios de la web.

No obstante, aún es posible afinar más y definir nuevas formas o metodologías de acceso a la información aún más específicas. Centrémonos por ejemplo en el caso de los usuarios de las bibliotecas digitales universitarias. Son usuarios muy especializados con unas necesidades de información muy definidas y específicas que por lo general tienen como objetivo encontrar recursos que les permitan profundizar en su dominio de conocimiento (accediendo por ejemplo a actas de congresos o artículos de revista actuales y novedosos sobre su área de especialización).

Así, cuando uno de estos de usuarios accede a un sistema de información que dispone de un servicio de filtrado y recomendaciones, la recomendación que se le ofrece se calcula conforme a la opinión que sobre un recurso tienen otros individuos con un perfil similar al del usuario activo (sistemas de recomendación cooperativos), o bien conforme a la similitud de un documento concreto con otros recursos ya valorados por el propio usuario (sistemas de recomendación basados en contenido).

En este trabajo hemos visto que para calcular esta semejanza existen múltiples medidas de similitud tradicionales como, por ejemplo, la función del *coseno de Salton* (Salton 1971) (Salton, Wong, Yang 1975); el *coeficiente de Dice* (van Rijsbergen 1979), o el *coeficiente de Jaccard* (Jaccard 1912) (Rorvig 1999), o medidas de similitud semánticas (ya comentadas en la sección 3.2.1.2). En todos estos sistemas a la hora de generar la recomendación, se hace una interpretación "*lineal*" de la función de similaridad utilizada. Es decir, a mayor similitud, mayor pertinencia del perfil (o del recurso) para calcular la recomendación. Esta interpretación, que denominamos "*monodisciplinar*", sería la más adecuada en la gran mayoría de los procesos de acceso a la información ya que permite orientar al usuario en su afán por obtener recursos que le

---

permitan profundizar en su área de especialización, desechando aquellos que no se corresponden con los criterios expresados en su perfil.

No obstante, en determinados casos es posible que esta forma de interpretar la función de similaridad no sea la más adecuada. Dada la naturaleza de los usuarios de las bibliotecas digitales universitarias no resulta infrecuente que muchos docentes e investigadores busquen ampliar sus líneas de investigación hacia dominios afines, o necesiten formar equipos de investigación o grupos de trabajo interdisciplinarios en los que converjan especialistas de áreas afines.

En estos casos para el investigador puede ser de gran utilidad descubrir recursos cuyo contenido sea reflejo de esa transversalidad entre dominios, disponer de la opinión que sobre los mismos tienen expertos de áreas distintas a la suya propia y, a ser posible, conocer la identidad de los expertos cuya valoración ha servido para generar dicha recomendación colaborativa, (y que posiblemente están desarrollando un trabajo con el que existe un cierto grado de conexión). Y todo ello sin que necesite modificar las preferencias ya definidas en su perfil.

En este caso particular no sería válida una interpretación lineal de la función de similaridad ya que el sistema debería considerar como más relevantes los recursos o perfiles cuya similaridad con el perfil del usuario activo tienen un valor "medio" (un valor entorno a 0.5).

Es decir, el sistema desechará aquellos recursos o perfiles cuyos valores de similaridad sean extremos con respecto al perfil del usuario activo (ya que esto implicaría respectivamente su pertenencia al dominio de especialización del usuario o a un área totalmente ajena).

En este caso, la función de similaridad no convendría interpretarla de una manera lineal sino que sería necesario definir algún tipo de función de centrado (Yager 2007) que permita delimitar el rango de valores de similaridad más adecuados para generar la recomendación. Para este caso concreto, la

---

interpretación de la similaridad podría estar definida por una función Gaussiana  $\mu$  como la siguiente,

$$\mu \left( Sim (p_i, r_j) \right) = e^{(Sim(p_i, r_j) - k)^2}$$

donde  $Sim (p_i, r_j)$  es la medida de similaridad entre el perfil de usuario  $p_i$  y el objeto  $r_j$  (que puede ser un documento u otro perfil de usuario), y  $k$  representa el valor central en torno al cual vamos a considerar que la similaridad es pertinente para generar la recomendación (en este caso  $k=0.5$ ).

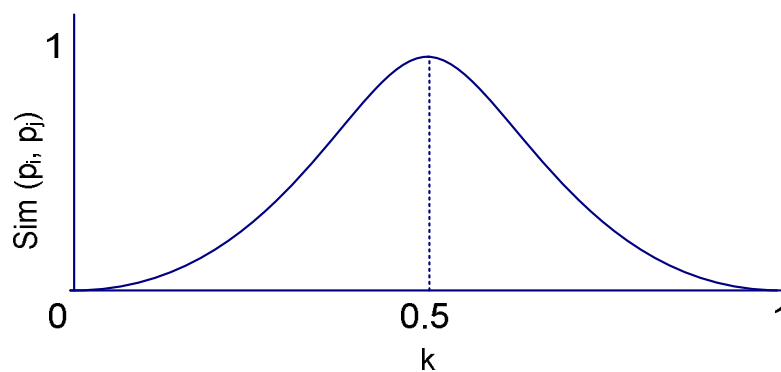


Fig.33 Representación de la función  $\mu$

De esta forma, un sistema de filtrado podría generar dos tipos de recomendaciones diferentes: la que denominamos *recomendación de especialización* o “*monodisciplinar*” que es aquella generada a raíz de una interpretación “lineal” de la función de similaridad entre perfiles, y otra que denominamos como *recomendación interdisciplinar* o *multidisciplinar*, que es generada a partir de la interpretación “no lineal” de la medida de similaridad, utilizando para ello una función de centrado.

Basado en este concepto, el prototipo D-Fusion se presenta como una aplicación que contempla ambos enfoques para proporcionar acceso personalizado a la información, abriendo de esta manera la puerta a nuevas

formas de entender la recuperación de información en un ámbito específico como las bibliotecas digitales universitarias.

## 5.2 ARQUITECTURA DE LA APLICACIÓN

D-Fussion es un prototipo de servicio de DSI diseñado para ser usado por usuarios de bibliotecas digitales universitarias, desarrollado aplicando tecnologías de Web Semántica y técnicas de modelado lingüístico difuso, y que permite obtener recomendaciones de especialización (“*monodisciplinarias*”) y recomendaciones transversales (multidisciplinarias).

El nombre del prototipo intenta plasmar el evidente carácter híbrido de la aplicación: un sistema cuyo objetivo es la difusión de información y que ha sido construido sobre vocabularios semánticos basados en RDF (RDF, RDFS, FOAF, RSS, etc.), y herramientas de lógica difusa (o *fuzzy*).

D-Fussion se ha desarrollado íntegramente con el lenguaje de programación de código abierto PHP 5.2.5 utilizando las librerías básicas del lenguaje y la librería DOM (específica para trabajar con documentos que utilizan sintaxis XML de manera que es posible tratarlos como árboles de nodos que pueden ser manipulados.), con la salvedad de una pequeña aplicación que permite desplegar el nombre de los usuarios cuyas valoraciones han sido utilizadas para generar recomendaciones colaborativas, que ha sido programada en JavaScript.

Los diferentes elementos con los que opera el sistema han sido definidos utilizando varios vocabularios de etiquetado basados en RDF, como RDF Schema, RSS 1.0 y FOAF. La idea de usar RDF como armazón sobre el que construir un sistema web frente a otros vocabularios más simples como XML se debe, no solo a las ventajas que ofrece la semántica, sino también a que usado de una manera pragmática, RDF es una herramienta que facilita la implementación de sistemas web gracias a su robustez.

---



### 5.3 ESTRUCTURA DEL PROTOTIPO

La estructura del prototipo es muy sencilla. Por un lado se han definido tres páginas web: una página de inicio (“*Demo*”) que contiene el formulario de entrada de datos donde es posible elegir el enfoque de recuperación preferido por el usuario (bien *monodisciplinar*, bien *multidisciplinar*), una página denominada “*Feedback*” donde se muestran los resultados al usuario, y una página de agradecimiento (“*Agradecimientos*”) que se genera cuando el usuario realiza una valoración alguno de los recursos devueltos por el sistema.

Por otro lado, se han creado varias librerías donde se definen los diferentes procedimientos y funciones que se utilizan en el programa para trabajar con el tesoro, los canales RSS y los perfiles de usuario. También se ha creado otra librería (“*Configuración*”) en la que es posible especificar y modificar diversos parámetros, como los umbrales de relevancia (tanto para la búsqueda *monodisciplinar* como para la *multidisciplinar*), o el número máximo de ítems relevantes que se van a mostrar a los usuarios.

Por último, la aplicación también dispone de dos directorios que funcionan a modo de repositorios: uno donde se almacenan los perfiles de los diferentes usuarios y otro que contiene canales RSS.

Para ejecutar la aplicación de manera local basta con tener instalado un servidor Apache 2.2.9 que soporte PHP 5.2.5, grabar la carpeta del prototipo en el directorio *root* y abrir el fichero “*demo.php*” con un navegador web.

---

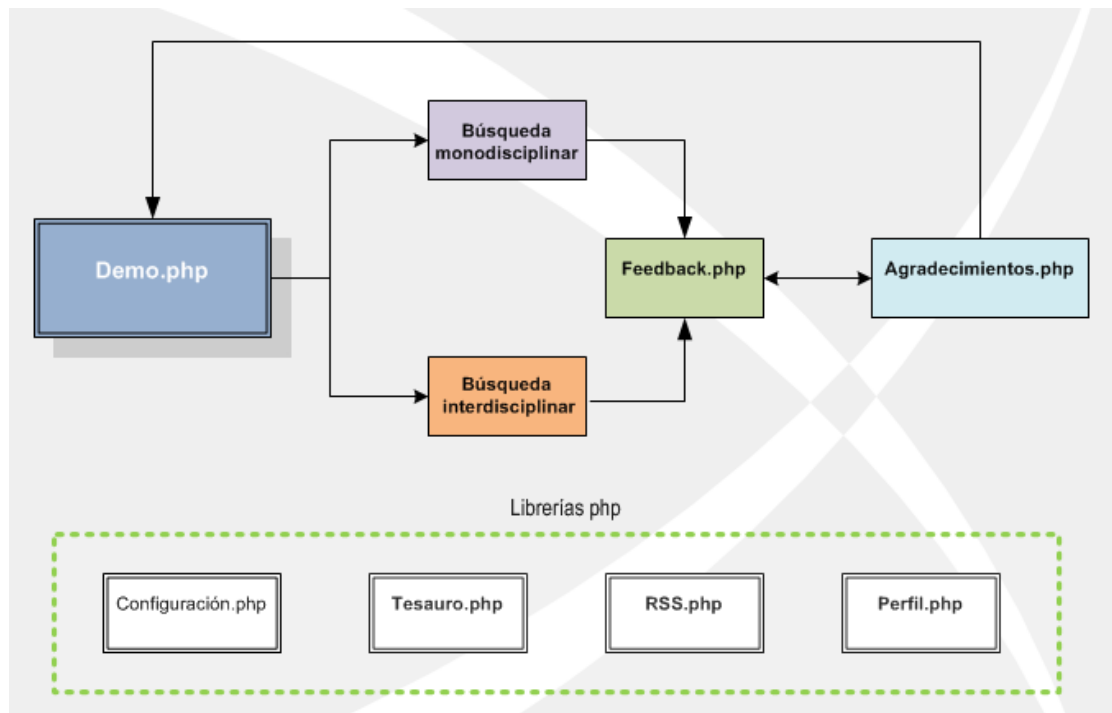


Fig.34 Estructura básica del prototipo D-fussion

### 5.3.1 ELEMENTOS

D-Fussion opera básicamente con tres elementos:

- *Tesauro:*

El tesauro sobre el que trabaja D-Fussion es un tesauro sencillo especializado en el área de Biblioteconomía y Documentación definido en formato RDFS. En nuestro caso, el tesauro utilizado es una versión del tesauro sobre «Biblioteconomía y Documentación» del CINDOC [172]. Contiene más de mil conceptos del área segmentados en 12 categorías temáticas principales o *top concepts* (*Biblioteconomía, Archivística, Museología, Fuentes de información, Profesionales de la información y usuarios, Estudios métricos de la información, Unidades de información, Ciencias y técnicas auxiliares, Tecnologías de la información y la comunicación, Lenguajes y lingüística, Sociedad de la información y Proceso documental*). En él nos limitamos a definir una serie clases y subclases relacionadas jerárquicamente que se corresponden con los

conceptos del dominio. Hemos prescindido de definir relaciones semánticas más complejas, como las asociativas, ya que no eran relevantes para la implementación de los módulos del prototipo. Por esta razón hemos utilizado RDF Schema para definir el tesoro en lugar de vocabularios más expresivos como SKOS Core (vocabulario que se define como extensión de RDF Schema y que está diseñado específicamente para desarrollar tesauros en la Web), ya que su expresividad es más que suficiente para desarrollar las funcionalidades del prototipo.

- *Perfiles de usuario:*

Los perfiles de usuario contienen una serie de datos básicos necesarios:

1. Una serie de datos de carácter personal (como nombre, información de contacto, etc.) que permiten identificar y caracterizar al usuario dentro del sistema.
  2. Datos de seguridad necesarios para que el sistema gestione el acceso al sistema del usuario (como un *login* y una contraseña).
  3. Una serie de preferencias relativas a sus intereses en el dominio de especialización de la biblioteca a las que debe asignar un peso o valor de relevancia que represente el grado de interés del usuario por esa materia en concreto. El número de términos de preferencia que formarán parte del perfil puede variar de acuerdo a las necesidades del usuario, aunque en nuestro caso hemos determinado un número máximo de tres preferencias por perfil.
  4. Un registro histórico de valoraciones sobre recursos que incluyen el identificador del recurso evaluado, la valoración asociada y la fecha en que se ha producido dicha valoración.
-

- *Canales RSS:*

La estructura de los canales RSS de acuerdo a la especificación del vocabulario es bastante sencilla ya que basta con indicar:

1. Los elementos básicos para la descripción del canal, como su título, un breve resumen de su contenido, datos relativos a su autoría y otra información específica para su procesamiento por parte de agentes software relativa a los periodos de actualización del mismo.

2. Los elementos básicos para la descripción de los ítems del canal. Los elementos mínimos necesarios para caracterizar los recursos son:

- Título del ítem.
  - Autores del documento.
  - Enlace al recurso: Permite al usuario acceder al texto completo del recurso. En nuestro sistema los documentos (generalmente en formato pdf) se encuentran en el repositorio de acceso libre E-LIS [11] que contiene recursos especializados del área de la Biblioteconomía y la Documentación.
  - Breve descripción del contenido.
  - Fecha de publicación.
  - Fuente original del recurso (referencia a la revista, libro, etc. de origen)
  - Una serie de conceptos que identifican las materias del ítem y que se han extraído del tesoro del sistema. Aunque el número de materias por documento puede variar recomendamos que este oscile entre 2 y 5 materias.
-

### 5.3.2 MÓDULOS

Se han implementado los principales módulos funcionales definidos en el modelo propuesto en el capítulo 4 (los módulos de *push de información*, de actualización de perfiles y de recomendación colaborativa) con la única excepción del módulo de creación de canales y perfiles, por lo que es preciso definir a priori los perfiles de usuario y los canales con los que el prototipo va a operar.

Además, D-Fussion permite elegir al usuario el enfoque con el que prefiere que el sistema genere las recomendaciones (*monodisciplinar* o multidisciplinar). Ambos enfoques solo se diferencian en dos aspectos:

- el criterio aplicado para ordenar los valores de relevancia (obtenidos en los módulos de *push de información* y en el de recomendación colaborativa) difiere dependiendo del enfoque.
- el módulo de actualización de perfiles no se aplica en el caso de las recomendaciones multidisciplinarias.

Como consecuencia, para un mismo perfil activo y canal RSS, y dependiendo del enfoque escogido, el sistema es capaz de generar diferentes conjuntos de recursos recomendados y grupos de expertos sin necesidad de modificar las preferencias del usuario.

#### 5.3.2.1 MÓDULO DE *PUSH DE INFORMACIÓN*

D-Fussion dispone de un interfaz a través del cual es posible elegir el perfil del usuario activo, el canal RSS, y el enfoque para generar las recomendaciones: *monodisciplinar* o multidisciplinar.

Una vez determinados los parámetros de entrada el sistema dispara la consulta al sistema y se equiparan las preferencias del perfil activo con las materias de los diferentes ítems almacenados en el canal RSS utilizando el operador taxonómico lineal definido en el capítulo 4, calculando la distancia que separa a los conceptos dentro de la estructura jerárquica del tesauro. Hemos optado por utilizar el

---

operador lineal ya que, aunque los resultados son algo menos precisos, los resultados son suficientemente representativos para comprobar la viabilidad del modelo.



The screenshot shows the 'Formulario inicial' (Initial Form) of the D-fussion service. The header features the 'D-fussion' logo and the text 'Servicio semántico-difuso de DSI' with a colorful cube icon. The form is titled 'Formulario inicial' and contains two main sections: 'Datos de entrada' (Input Data) and 'Acciones' (Actions). In the 'Datos de entrada' section, there are two dropdown menus: 'Canal' (Channel) set to 'doc2.rss' and 'Perfil' (Profile) set to 'perfil.rdf'. The 'Acciones' section has two radio buttons: 'Búsqueda monodisciplinar' (Monodisciplinary Search) which is selected, and 'Búsqueda multidisciplinar' (Multidisciplinary Search). Below these sections is an 'Aceptar' (Accept) button. At the bottom of the form, there is a footer with the text: 'Para cualquier comentario ponerse en contacto con el [administrador](#). Copyright 2005-2008. Última actualización: 19 Junio 2008'. At the very bottom, there are two accessibility logos: 'W3C WAI-AA WCAG 1.0' and 'W3C HTML 4.01' with a red checkmark.

Fig.35 Interfaz de entrada de datos

Posteriormente se obtiene la relevancia entre perfil e ítems utilizando la función de *solapamiento semántico*. Dependiendo del enfoque definido para obtener las recomendaciones el sistema procede a ordenar los recursos de acuerdo a este valor de relevancia para presentarlos en pantalla.

Para evitar una sobre carga de información (hay que tener en cuenta que los canales RSS contienen hasta 15 ítems) se define un umbral mínimo de relevancia para evitar la presentación de ítems con una relevancia poco representativa. Este umbral mínimo de relevancia es  $k=0.50$  en el caso del enfoque *monodisciplinar*, y  $k=1.09$  para el *multidisciplinar* (siendo en ambos casos la relevancia máxima igual a 1). Este valor de umbral puede ser modificado de acuerdo a las necesidades del sistema para relajar el nivel de restricción a conveniencia, pudiendo ser aumentado cuando el grado de silencio documental es demasiado elevado.

El sistema define un nivel adicional de filtrado limitando el número de ítems relevantes mostrados al usuario para evitar la sobrecarga de información (hay que tener en cuenta que en el canal RSS se almacenan hasta 15 ítems diferentes y el objetivo del sistema es ofrecer una alerta que el usuario pueda revisar de un vistazo). Presentando un número de ítems no mayor a 5 garantizamos que el usuario dispone de un número razonable de ítems para examinar y facilitamos que se anime a evaluarlos.

The screenshot shows the 'D-fussion' search results page. The header includes the logo and the text 'Servicio semántico-difuso de DSI'. Below the header, the page is titled 'Resultados de la búsqueda'. The results are presented in a table-like format with four columns: 'Documento', 'Relevancia', 'Recomendación colaborativa', and 'Valoración'.

Documento	Relevancia	Recomendación colaborativa	Valoración
<p><b>El cambio organizacional</b></p> <p>En ésta presentación se enuncian los principales conceptos del cambio organizacional: cambios en las personas, resistencia al cambio; cambios no planificados; cambios planificados; cambios impuestos; cambios participativos; cambios negociados. El objetivo del Seminario es inducir a los participantes a aceptar la necesidad de cambiar los procesos y la cultura de la biblioteca.</p>	Muy alta +0.000	<p>Media</p> <ul style="list-style-type: none"> <li>Francoise Daguerre</li> </ul>	<p>Satisfacción:</p> <p>Nula</p> <p>Valorar</p>
<p><b>Webs sempre accessibles : les biblioteques nacionals i els dipòsits digitals nacionals = Webs sempre accessibles : las bibliotecas nacionales y los depósitos digitales nacionales</b></p> <p>Las tecnologías de la información y la comunicación han facilitado que el patrimonio cultural, científico y la información en general se presenten en formato digital, así como en los formatos analógicos tradicionales. La reacción no se ha hecho esperar, y desde la década de los noventa han surgido diversos proyectos destinados a garantizar el acceso permanente a la producción digital -su recopilación y almacenaje, el tratamiento, la preservación y la difusión-. Se presenta la panorámica mundial de los modelos existentes de depósitos digitales nacionales, nombre que reciben estos proyectos impulsados habitualmente por las bibliotecas nacionales, con un objetivo común: hacer que las páginas web sean siempre accesibles.</p>	Media +0.081		<p>Satisfacción:</p> <p>Nula</p> <p>Valorar</p>

Fig.36 Relevancia de los resultados

Los recursos aparecen ordenados por orden de relevancia y de cada uno de ellos se muestra título, resumen, enlace al documento original a texto completo, y una etiqueta lingüística con un modificador numérico que representa el valor de relevancia (obtenida de acuerdo al modelado lingüístico difuso basado en 2-tuplas).

Junto a este valor también aparece una recomendación colaborativa (en la siguiente sección explicamos cómo se calcula) y un menú desplegable a través del cual al usuario puede valorar cada ítem.

### 5.3.2.2 MÓDULO DE RECOMENDACIÓN COLABORATIVA

Este módulo lo que pretende es aprovechar la experiencia de los expertos que forman parte de la comunidad de usuarios de la biblioteca.

Para ello, junto al valor de relevancia de los ítems del canal RSS (que se calcula de acuerdo a las preferencias que tiene definidas el usuario en su perfil), el sistema ofrece otro elemento que le permite al usuario disponer de más información sobre la calidad de los documentos recomendados por el sistema. Este nuevo elemento es una recomendación colaborativa que se calcula a partir de las valoraciones que sobre los recursos emiten usuarios con un perfil similar al del usuario activo.

El proceso consta de 3 etapas:

- En una primera etapa el sistema calcula la similaridad de los perfiles almacenados en el repositorio de perfiles del sistema con el perfil del usuario activo. El cálculo es análogo al que se realiza cuando se compara un perfil con un ítem del canal RSS. El resultado es un *cluster* de usuarios afines.
  - La segunda etapa consiste en comprobar si los recursos recomendados por el sistema en la fase de *push* de información han sido evaluados por alguno de los usuarios que forman parte del *cluster*.
  - Por último en la tercera etapa, el sistema procede a agregar las valoraciones de los diferentes expertos del *cluster* de perfiles afines para cada recurso utilizando el operador lingüístico LOWA, generando una recomendación colaborativa (que denominamos *monodisciplinar*) para cada uno de ellos. Este operador permite la agregación de etiquetas lingüísticas con igual nivel de importancia y su salida es una nueva etiqueta lingüística perteneciente a la variable lingüística “*Nivel de satisfacción*”.
-



A la hora de presentar la información en pantalla el sistema permite al usuario hacer *click* sobre la etiqueta lingüística que representa la recomendación colaborativa para desplegar un listado con los nombres de los expertos cuyas valoraciones han servido para generar dicha recomendación.

**D-fusion** Servicio semántico-difuso de DSI

Resultados de la búsqueda

Documento	Relevancia	Recomendación colaborativa	Valoración
<p><b>El cambio organizacional</b></p> <p>En ésta presentación se enuncian los principales conceptos del cambio organizacional: cambios en las personas, resistencia al cambio; cambios no planificados; cambios planificados; cambios impuestos; cambios participativos; cambios negociados. El objetivo del Seminario es inducir a los participantes a aceptar la necesidad de cambiar los procesos y la cultura de la biblioteca.</p>	Muy alta +0.000	<p>Media</p> <ul style="list-style-type: none"> <li>Francoise Daguerre</li> </ul>	<p>Satisfacción:</p> <p>Nula</p> <p>Valorar</p>
<p><b>Webs siempre accesibles : les biblioteques nacionals i els dipòsits digitals nacionals = Webs siempre accesibles : las bibliotecas nacionales y los depósitos digitales nacionales</b></p> <p>Las tecnologías de la información y la comunicación han facilitado que el patrimonio cultural, científico y la información en general se presenten en formato digital, así como en los formatos analógicos tradicionales. La reacción no se ha hecho esperar, y desde la década de los noventa han surgido diversos proyectos destinados a garantizar el acceso permanente a la producción digital -su recopilación y almacenaje, el tratamiento, la preservación y la difusión-. Se presenta la panorámica mundial de los modelos existentes de depósitos digitales nacionales, nombre que reciben estos proyectos impulsados habitualmente por las bibliotecas nacionales, con un objetivo común: hacer que las páginas web sean siempre accesibles.</p>	Media +0.081		<p>Satisfacción:</p> <p>Nula</p> <p>Valorar</p>

Fig.37 Recomendación colaborativa

Este listado de nombres supone una extraordinaria fuente de información adicional para el usuario ya que, por un lado, puede determinar el nivel de confianza que le merece la recomendación generada por el sistema (dependiendo de si los expertos son profesionales de prestigio con una opinión autorizada), y por otro lado puede descubrir a otros profesionales expertos en su misma área de interés con los que poder establecer grupos de trabajo o emprender proyectos de investigación conjuntos.

### 5.3.2.3 MÓDULO DE ACTUALIZACIÓN DE PERFILES O *FEEDBACK*

Como resultado del proceso de *push* de información el usuario recibe el listado de los ítems del canal RSS que pueden resultarle de interés, y tiene posibilidad de examinarlos documentos sugeridos bien a través del resumen del contenido de los mismos o bien accediendo al texto completo a través del enlace que da acceso al mismo.

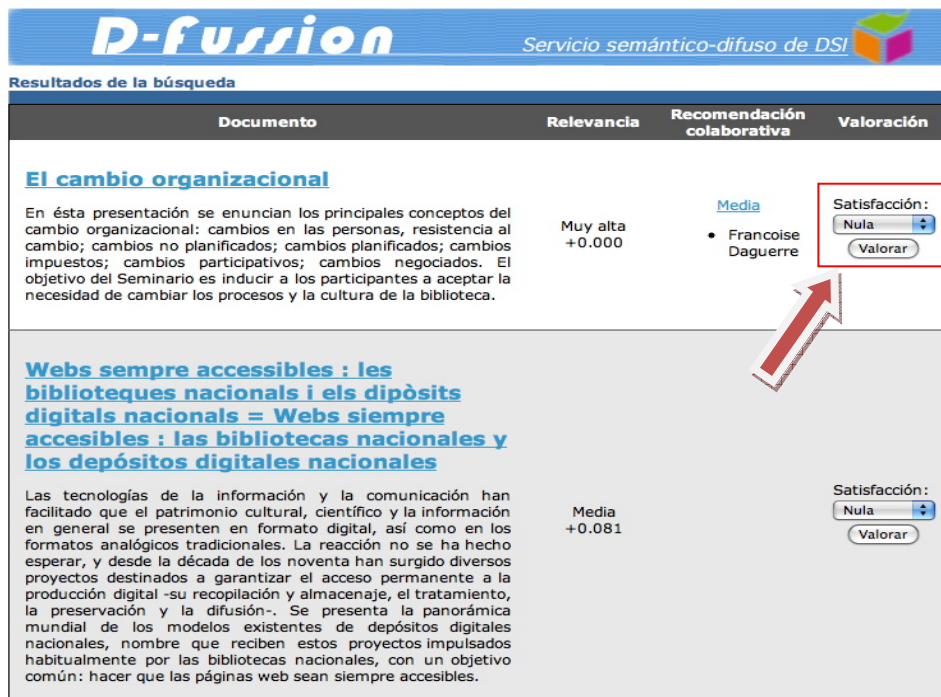


Fig.38 Menú desplegable de valoración

Una vez examinados los recursos, el sistema da al usuario la opción de evaluar estos documentos utilizando el desplegable que se encuentra bajo cada ítem. En este desplegable aparece un conjunto de etiquetas lingüísticas “*Nivel de satisfacción*” que permiten al usuario emitir su juicio con respecto a la idoneidad del recurso recomendado por el sistema.

Esta valoración es de vital importancia para el sistema ya que cumple un doble objetivo. Por un lado es almacenada en el perfil del usuario en el registro histórico de valoraciones (que es utilizado por el sistema para generar las

recomendaciones colaborativas), y por otro lado sirve para disparar el mecanismo de actualización de perfiles.

El mecanismo de actualización funciona de la siguiente manera. Cuando el usuario define su valoración sobre el recurso el sistema determina cuál es la preferencia del perfil que más similaridad presenta con las materias del ítem. El peso asociado a dicha preferencia será el peso a actualizar mediante la función de actualización definida en el capítulo 4.

Este mecanismo permite identificar de forma tácita las variaciones en las preferencias de los usuarios ya que se procede a bonificar o penalizar las preferencias que han tenido un mayor peso a la hora de calcular la relevancia del recurso que se está valorando. De esta manera, el hecho de que un usuario, repetidamente en el tiempo, valore de forma negativa documentos donde una preferencia determinada de su perfil tiene un mayor peso específico puede indicar una variación significativa en sus preferencias. En este caso el sistema deberá contar con medios necesarios para detectar esta circunstancia y alertar al usuario de esta circunstancia para que pueda redefinir las preferencias de su perfil y así caracterizar de una manera más precisa sus necesidades de información.

La idea que subyace al mecanismo de actualización es que el sistema supere el modelo de recuperación de información basada en perfiles estáticos y sea capaz de adaptarse de forma dinámica a las variaciones en las preferencias de los usuarios, pero sin perder la perspectiva de que el perfil debe reflejar las preferencias del usuario de una manera estable en el tiempo a medio/largo plazo.

Por esta razón, y aunque en nuestro prototipo la actualización del perfil se produce cada vez que el usuario evalúa un recurso, lo ideal sería definir políticas de actualización que realizaran el recálculo de pesos de forma periódica

---

(estableciendo plazos de semanas o meses), o bien disparando el proceso en el momento en que el usuario acumula un número determinado de valoraciones.

Este módulo no está disponible cuando se activa el enfoque multidisciplinar para generar las recomendaciones ya que, en este caso, no tiene sentido bonificar o penalizar preferencias que en el fondo no representan fielmente los intereses reales del usuario.

## 5.6 EJEMPLO OPERACIONAL

Supongamos un usuario genérico que desea recibir recomendaciones *monodisciplinares* y del cual el sistema mantiene un perfil  $P$  donde se definen las preferencias  $\alpha_1, \alpha_2$ , cuyos pesos asociados son  $\omega_1, \omega_2$ .

Supongamos también un ítem  $R$  del canal RSS de la biblioteca cuyo contenido viene representado por las materias  $\beta_1, \beta_2, \beta_3$ .

Para calcular la similaridad entre conceptos dentro de la taxonomía definida en el tesauro aplicamos la función definida por Oldakowsky y Bizer (2005) utilizando el operador lineal para calcular la distancia entre conceptos.

Sea  $\alpha_1$  el concepto "*Instrumentos de control*" con una profundidad en el tesauro de 2, y  $\beta_2$  "*Clasificación de fondos*" con una profundidad en el árbol de 3 (siendo la profundidad total del tesauro de 6), siendo "*Archivística*" el nodo padre común más cercano (*ccp*) cuya profundidad en el tesauro es 0 ya que es un término cabecera de clase.

---

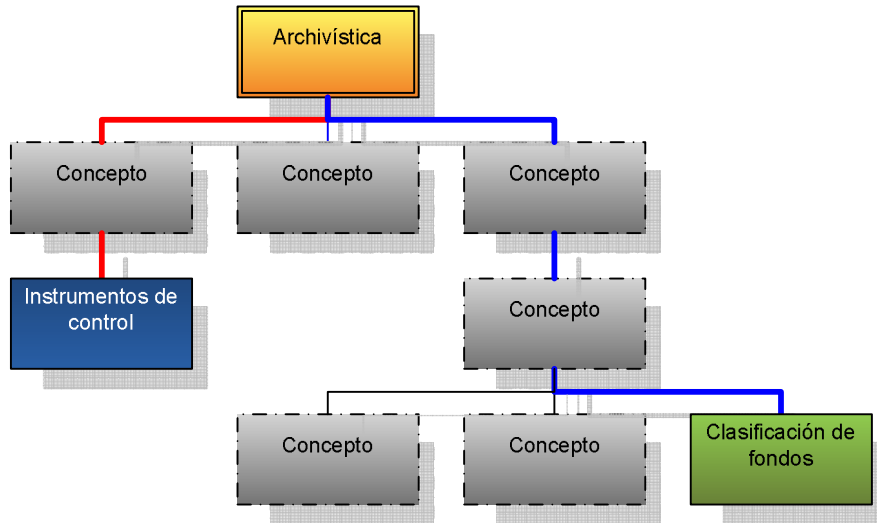


Fig.39 Situación de conceptos en el tesaurus

Aplicando la función lineal de cálculo de *hitos* obtenemos:

$$milestone (\alpha_1)= 0.67 ; milestone (\beta_2)= 0.50 ; milestone (ccp)=1$$

por lo que la distancia de cada concepto al *ccp* es:

$$d(\alpha_1, ccp)= 0.33 \text{ y } d(\beta_2, ccp)= 0.50$$

y por lo tanto, la distancia entre  $\alpha_1$  y  $\beta_2$  es:

$$d (\alpha_1, \beta_2)= 0.83$$

Sean el resto de distancias las que se presentan en la tabla 5, obtenemos las similaridades de la tabla 6 :

Preferencias/Materias	$\beta_1$	$\beta_2$	$\beta_3$
$\alpha_1$	0.21	0.03	0.35
$\alpha_2$	0.16	0.07	0.12

Tabla 5. Distancias

Preferencias/Materias	$\beta_1$	$\beta_2$	$\beta_3$
$\alpha_1$	0.79	0.97	0.65
$\alpha_2$	0.84	0.93	0.88

Tabla 6. Similaridades

A continuación se calcula la relevancia del ítem  $R$  con el perfil  $P$  de acuerdo a la función:

$$Sim(P_i, R_j) = \frac{\sum_{k=1}^{MIN(N,M)} H_k (Sim(p_i, r_j)) \omega_i}{MAX(N, M)}$$

Supongamos que para el usuario el peso o importancia de la preferencia  $\alpha_1$  es “Muy alta” cuya correspondencia numérica de acuerdo a la semántica de la variable lingüística “Nivel de relevancia” es  $\omega_1 = 0.83$ , y para  $\alpha_2$  es “Muy baja”, o lo que es lo mismo  $\omega_2 = 0.5$ .

Dado que  $N=M=3$ , las 3 similaridades máximas son  $Sim(\alpha_1, \beta_2)$ ,  $Sim(\alpha_2, \beta_2)$ , y  $Sim(\alpha_2, \beta_3)$ , por lo que obtenemos la siguiente relevancia del recurso  $R$  con respecto al perfil  $P$ :

$$Rel(P, R) = 0.57$$

Dado que el umbral de relevancia está fijado en  $k=0.50$  este recurso sería presentado al usuario ordenado de acuerdo a este valor.

Aplicando el enfoque de modelado lingüístico difuso basado en 2-tuplas la relevancia se presenta al usuario como una etiqueta lingüística extraída de la variable lingüística “Nivel de relevancia” y un valor numérico de la siguiente forma:

$$“Media” + 0.07$$

ya que la etiqueta lingüística más próxima a 0.57 se corresponde con el valor “Media”.

A continuación, el sistema busca perfiles similares al del usuario que hayan evaluado el recurso R para generar la recomendación colaborativa para dicho recurso. Suponiendo que se han encontrado dos perfiles afine P' y P'' que han valorado el recurso con las etiquetas "Alta" y "Media" extraídas de la variable lingüística "Nivel de satisfacción".

Aplicando el operador LOWA (con  $\omega_1 = 0.4$ ) la etiqueta resultante de la agregación es:

$$k = \text{MIN}\{6, 3 + \text{round}(0.4 * (4 - 3))\} = 3 \Rightarrow l_k = \text{Media}$$

Dado que la similaridad media no ponderada de  $\alpha_1$  (tiene un valor de 0.80) es menor que la de  $\alpha_2$  (con un valor de 0.88), esta última sería la actualizada de acuerdo a la evaluación del usuario. Veamos un ejemplo de cómo se actualizaría un perfil.

Supongamos que el usuario decide evaluar el recurso  $\beta_2$  recomendado por el sistema, y considera que satisface en gran medida sus necesidades de información y pondera su grado de satisfacción con la etiqueta lingüística  $e_j = \text{«alta»}$  (donde  $e_j \in S' = \{nula, muy\ baja, baja, media, alta, muy\ alta, total\}$ ). Entonces el sistema procede a actualizar las frecuencias de las áreas de interés que están presentes entre las materias del recurso evaluado. En este caso el área de interés «Archivística» tiene asociado un peso  $\omega^j_{(Preferencia, Archivística)} = \text{«muy bajo»}$  (donde  $\omega^j_{li} \in S = \{nulo, muy\ bajo, bajo, medio, alto, muy\ alto, total\}$ ). Considerando que  $s_a \leq s_b$ , que los valores de los índices  $a$  y  $b$  son 1 y 4 respectivamente, y que  $T=6$ , tenemos que  $\beta=1$ , el nuevo peso para este valor de preferencia se incrementa en un factor de 1, y por lo tanto  $(\omega^j_{(Preferencia, Archivística)})' = g(\text{alta, muy bajo}) = \text{«baja»}$ .

En el caso de que el usuario hubiera decidido recibir recomendaciones multidisciplinares el proceso se desarrolla de una manera ligeramente diferente. Suponiendo que el usuario ha recibido como resultado de la consulta pasiva los documentos R y R', con unos valores de relevancia con  $Rel(P, R)=0.57$  y  $Rel(P,$

$R'=0.83$ . El sistema procede a ordenar ambos recursos de acuerdo a su nivel de relevancia utilizando la función de centrado  $\mu$ :

$$\mu(\text{Rel}(P, R)) = 1.005$$

$$\mu(\text{Rel}(P, R')) = 1.110$$

El sistema procede a reordenar los documentos de acuerdo a estos nuevos valores, considerando más relevantes los valores más cercanos a 1 (en este caso  $R$  sería más relevante que  $R'$  y a la hora de presentarlo en pantalla al usuario aparecería por encima).

El procedimiento sería análogo para calcular la semejanza del perfil  $P$  con los perfiles del resto de expertos.

---





---

## CAPÍTULO 6

# RESULTADOS, CONCLUSIONES Y TRABAJOS FUTUROS

### 6.1 RESULTADOS

El objetivo que hemos perseguido con este trabajo ha sido estudiar el desarrollo de sistemas de filtrado y recomendación que permitan mejorar el acceso a recursos informativos de interés a los usuarios de las bibliotecas digitales universitarias. Para conseguirlo, hemos propuesto combinar el uso de tecnologías de Web Semántica y técnicas de modelado lingüístico difuso para i) mejorar la representación de la información y la comunicación entre usuario y sistema, ii) mejorar de la satisfacción de las necesidades de información del usuario mediante el filtrado de recursos de acuerdo a las preferencias e intereses recogidos en su perfil, y iii) desarrollar nuevos servicios que permitan satisfacer necesidades de información específicas de la comunidad de usuarios a la que sirven las bibliotecas digitales universitarias (como la búsqueda monodisciplinar y la multidisciplinar).

Atendiendo a estos aspectos, los resultados obtenidos en esta memoria de tesis pueden resumirse en los siguientes puntos:

- Hemos analizado la Web Semántica como nuevo paradigma tecnológico que está afianzándose como un modelo alternativo (y viable) al actual modelo de Web.
  - También se ha detallado el papel que las bibliotecas digitales universitarias van a jugar en el nuevo marco de Enseñanza Superior Europea como elemento dinamizador de la investigación y la
-

docencia, y la necesidad de desarrollar nuevos servicios que satisfagan los requerimientos cada vez más exigentes de los usuarios.

- Hemos estudiado algunos de los servicios de valor añadido que es posible implementar en una biblioteca digital, en concreto la redifusión de contenidos y los sistemas de filtrado y recomendación, así como algunas de las tecnologías y herramientas necesarias para ello: los sistemas de organización del conocimiento, las tecnologías de Web Semántica y las técnicas de modelado lingüístico difuso.
- Hemos propuesto un modelo de servicio de difusión selectiva de información para bibliotecas digitales basado en tecnologías de Web Semántico, técnicas de filtrado de información y técnicas de modelado lingüístico difuso construido sobre una plataforma multi-agente.
- Se ha diseñado e implementado un prototipo de servicio de DSI para bibliotecas digitales que utiliza conjuntamente tecnologías de Web Semántica, técnicas de filtrado de información colaborativas y basadas en contenido, y técnicas de modelado lingüístico (tanto ordinal como basado en 2-tuplas).

Algunos de los desarrollos que conforman esta memoria han sido ya publicados en (Peis et al. 2006) (Herrera-Viedma et al. 2007) (Morales-del-Castillo, Herrera-Viedma, Peis 2007) (Peis, Morales-del-Castillo, Herrera-Viedma 2008) (Peis, Morales-del-Castillo, Delgado-López 2008).

## **6.2 CONCLUSIONES**

En el capítulo anterior hemos presentado un servicio de difusión selectiva de información y recomendaciones cuyo propósito es facilitar el acceso a los usuarios de una biblioteca digital universitaria las novedades bibliográficas que más se adecuan a sus intereses.

---

El sistema presenta diversos elementos que añaden valor a la información que se ofrece a los usuarios:

- El uso de las tecnologías de Web Semántica permiten definir descripciones enriquecidas de los recursos, haciéndolos semánticamente accesibles, interoperables y fácilmente interpretables tanto por los usuarios como por el sistema. La utilización de estos vocabularios deja la puerta abierta a la ampliación del servicio de filtrado y recomendación con nuevas funcionalidades que permitieran, por ejemplo, definir nuevos criterios de filtrado (como el nivel de conocimientos del usuario o la tipología documental de los recursos), u obtener recursos relacionados con las materias de interés de los usuarios aprovechando las capacidades de inferencia que ofrecería una infraestructura ontológica adecuada.
  - La redifusión de contenidos permite ofrecer al usuario información actualizada de una manera rápida gracias a la facilidad de edición de los canales RSS. Además, el uso de estas herramientas permite no solo ofrecer información a los usuarios de la biblioteca, sino que usados en su forma habitual pueden servir como un canal para difundir los contenidos de la biblioteca al resto de usuarios de la Web.
  - Las técnicas de modelado lingüístico (tanto ordinal como 2-tuplas) permiten que las recomendaciones generadas por el sistema sean más comprensibles para el usuario y de esta forma disponga de más elementos de juicio para decidir sobre la idoneidad de los recursos recomendados por el sistema.
  - El servicio ofrece varios niveles de filtrado basado en contenido, imponiendo restricciones sobre el grado mínimo de relevancia de los recursos con respecto a las preferencias del usuario, y sobre el número de recursos interesantes máximo que se presentan al usuario (permitiendo de esta forma que el número de ítems recomendados no lo abrumen y disponga de un número razonable que pueda
-

examinar de una manera rápida). Esta circunstancia creemos que además facilita que el usuario se anime a evaluar los recursos recomendados, colaborando de esta forma activamente a mantener el sistema vivo y actualizado.

- El sistema es capaz de generar recomendaciones colaborativas orientadas a la especialización en el área de investigación del usuario (recomendaciones de especialización o “monodisciplinares”), y recomendaciones interdisciplinares que conectan su área de investigación con otras disciplinas. Mientras que la recomendación de especialización supone una interpretación “lineal” de la función de similaridad utilizada para equiparar los perfiles de usuarios, la recomendación interdisciplinar se genera gracias al uso de una función de centrado que permite reinterpretar la medida de similaridad. De esta forma al investigador no solo se le ofrecen recomendaciones sobre recursos pertinentes a sus necesidades sino que además puede descubrir redes sociales implícitas que lo interrelacionan con otros profesionales (tanto de su área de conocimiento como de otros dominios) con los que poder crear grupos de trabajo.
  - La recomendación generada como producto del filtrado colaborativo es una rica fuente de información para el usuario que puede resultarle interesante desde tres vertientes diferentes:
    - Por un lado, el usuario dispone de una etiqueta lingüística que representa la recomendación generada a partir de las opiniones de otros usuarios con intereses similares, lo cual le puede orientar sobre la idoneidad de un recurso.
    - Por otro lado, al dar el sistema la opción de conocer el nombre de los expertos cuyas valoraciones han servido para generar la recomendación colaborativa el usuario puede determinar, según su criterio, la confianza (o fiabilidad) que le merece
-

dicha recomendación (definiendo de esta forma una especie de filtrado basado en confianza).

- Por último, gracias a que el sistema soporta tanto el enfoque *monodisciplinar* como el multidisciplinar a la hora de generar las alertas, y a que es posible conocer el nombre de los expertos cuyas valoraciones han servido para generar la recomendación colaborativa, el usuario puede descubrir redes sociales implícitas (bien de su área o de áreas afines) con cuyos miembros puede establecer relaciones profesionales de colaboración.

### 6.3 TRABAJOS FUTUROS

En primer lugar queremos apuntar algunas funcionalidades que sería deseable incluir para mejorar el prototipo que presentamos en esta memoria:

- Sería deseable integrar en los módulos de creación de perfiles y de canales RSS un sistema que permitiera navegar por el tesoro de manera que, tanto usuario como administrador, pudieran seleccionar respectivamente las preferencias del perfil y las materias de los ítems del canal cuando las que sugiere el sistema no resultan satisfactorias.
  - Otro servicio deseable consistiría en permitir al usuario obtener la referencia bibliográfica completa de aquellos ítems que le han resultado interesantes de manera que pueda aprovechar esta información en sus trabajos o investigaciones.
  - Sería imprescindible definir mecanismos de fusión de ontologías (o esquemas de conceptos) para aumentar la cobertura temática del servicio, ya que este prototipo se restringe exclusivamente a la generación de alertas y recomendaciones en el campo de la Biblioteconomía y la Documentación (lo cual limita sobre todo el enfoque de búsqueda multidisciplinar).
-

- Dado que los perfiles de usuario están descritos con FOAF sería muy simple definir de forma explícita redes sociales (mediante el elemento *foaf:knows*) lo cual facilitaría aún más la tarea de descubrimiento de usuarios afines y aumentarían las posibilidades de crear nuevos grupos de investigación (tanto *monodisciplinares* como multidisciplinares).
- Otra funcionalidad interesante consistiría en implementar un mecanismo de recomendaciones colaborativas basado en la tipología o el nivel de conocimientos de los usuarios.

También vamos a destacar en este apartado algunas de las líneas de investigación que a nuestro parecer resultan especialmente interesantes de cara a futuros trabajos en el desarrollo de sistemas de difusión selectiva de información.

- Avanzar en el estudio de la integración de las técnicas de modelado lingüístico con las tecnologías de Web Semántica y su utilización en diferentes dominios de aplicación, ya que el uso conjunto de ambas supone una potente herramienta que facilita y agiliza la comunicación del usuario de la Web con los agentes software.
  - Aprovechar el potencial de nuevos vocabularios semánticos como RDFa, eRDF y GRDDL para embeber las descripciones semánticas de los recursos directamente en el código (X)HTML, pudiendo de esta manera operar directamente con los documentos originales y no con representaciones de estos.
  - Desarrollar mecanismos más complejos de actualización de perfiles que permitan al sistema “aprender” del comportamiento del usuario para ofrecer resultados más adecuados a sus necesidades. incluyan la monitorización del comportamiento de los usuarios y técnicas de minería web semántica de manera que sea posible capturar y actualizar las preferencias del usuario de una manera más sencilla y precisa.
-

- Profundizar en el estudio de las semánticas emergentes que se encuentran implícitas en las comunidades cerradas de usuarios para generar, mantener y actualizar ontologías (especialmente en el ámbito de las bibliotecas digitales universitarias).
  - Estudiar la posibilidad de desarrollar mecanismos de construcción semi-automática de tesauros y ontologías utilizando técnicas de procesamiento del lenguaje natural (PLN) e integrarlos en sistemas de filtrado similares al que presentamos en este trabajo.
  - Por último, otra línea de investigación muy interesante consiste en el desarrollo de sistemas y redes científicas de información que dispongan de herramientas de filtrado basadas en redes de confianza, de manera que se facilite a los investigadores descubrir instituciones, publicaciones o personas autorizadas en un dominio de conocimiento determinado.
-





---

## BIBLIOGRAFÍA

1. Adida, B. y Birbeck, M. (2007) "RDFa Primer: Embedding Structured Data in Web Pages." En línea, [consultado 28 febrero 2008]. Disponible en <http://www.w3.org/TR/xhtml-rdfa-primer/>.
  2. Adrian, B.; Sauermann, L. y Roth-Berghofer, T. (2007) "ConTag: A Semantic Tag Recommendation System". En: Proceedings of I-MEDIA' 07 and I-SEMANTICS' 07 International Conferences on New Media Technology and Semantic Systems as Part of (TRIPLE-I-2007), pp. 297-304.
  3. Aduna, B. V. y Sirma, A., (eds.) (2005) "The SERQL query language (revision 1.2)." En línea, [consultado 5 octubre 2007]. Disponible en <http://www.openrdf.org/doc/sesame/users/ch06.html>.
  4. Aksoy, D. et al. (1998) "Research in Data Broadcast and Dissemination". En: Proc. Of the 1<sup>st</sup> International Conference on Advanced Multimedia Content Processing, pp. 194-207.
  5. Alonso, G. et al. (2004) "Web Services: Concepts, architectures and applications". Berlin: Springer Verlag.
  6. Altinel, M. y Franklin, M.J. (2000) "Efficient Filtering of XML Documents for Selective Dissemination of Information". En: Proc. of the 26th VLDB Conference. pp. 53-64.
  7. Antinou, G. y van Harmelen, F. (2004A). "A Semantic Web Primer". MIT Press.
  8. Antinou, G. y van Harmelen, F. (2004B) "Web Ontology Language: OWL". En: S. Staab, y R. Studer (eds.) Handbook on ontologies, pp. 67-92. Berlin: Springer Verlag.
  9. Arano, S. y Codina, L. (2004) "La estructura conceptual de los tesauros en el entorno digital: ¿nuevas esperanzas para viejos problemas?" 9es Jornades catalanes d'informació i documentació. En línea, [consultado 10 enero 2008]. Disponible en <http://www.lluiscodina.com/ontotesauros.doc>.
  10. Balabanovic, M. y Shoham, Y. (1997) "Fab: content-based, collaborative recommendation". Communications of the ACM 40, pp. 66-72.
  11. Barrueco, J. M. (2007) "Desarrollo de servicios de valor añadido". En: Actas del congreso La ecología de los repositorios institucionales. En línea, [consultado 12 abril 2008]. Disponible en <http://redsicura.iata.csic.es/xarxa/ocs/papers/Gijon-Barrueco.ppt>.
  12. Bartel, M. et al. (2002) "XML signature syntax and processing." En línea, [consultado 20 agosto 2007]. Disponible en <http://www.w3.org/TR/xmlsig-core/>.
  13. Basu, C.; Hirsh, H. y Cohen, W. (1998) "Recommendation as classification: Using social and content-based information in recommendation". En: Proceedings of the 15th National Conference on Artificial Intelligence, pp. 714-720.
-

14. Battle, S. et al. (2005) "Semantic Web Services Language (SWSL)." En línea, [consultado 1 noviembre 2007]. Disponible en <http://www.w3.org/Submission/SWSF-SWSL/>.
  15. Beckett, D., (ed.) (2004) "RDF/XML Syntax Specification (Revised)." En línea, [consultado 13 febrero 2007]. Disponible en <http://www.w3.org/TR/rdf-syntax-grammar/>.
  16. Bedi, P.; Kaur, H. y Marwaha, S. (2007) "Trust based Recommender System for the Semantic Web". En: Proceedings of the IJCAI 07, pp. 2677-2682.
  17. Beged-Dov, G. et al. (2000) "RDF Site Summary 1.0 Modules: Dublin Core." En línea, [consultado 6 enero 2007]. Disponible en <http://web.resource.org/rss/1.0/modules/dc/>.
  18. Beged-Dov, G. et al. (2001A) "RDF Site Summary 1.0 Modules: Syndication." En línea, [consultado 21 junio 2007]. Disponible en <http://web.resource.org/rss/1.0/modules/syndication/>.
  19. Beged-Dov, G. et al. (2001B) "RDF Site Summary (RSS) 1.0." En línea, [consultado 6 septiembre 2007]. Disponible en <http://web.resource.org/rss/1.0/spec>.
  20. Benjamins, R. et al. (1999) "(KA)2: building ontologies for the Internet: a midterm report". International Journal of Human-Computer Studies 51 (3), pp. 687-712.
  21. Berglund, A., (ed.) (2006) "eXtensible Stylesheet Language (XSL) Version 1.1." En línea, [consultado 14 febrero 2008]. Disponible en <http://www.w3.org/TR/xsl/>.
  22. Berglund, A. et al., (ed.) (2007) "XML Path Language (XPath) 2.0." En línea, [consultado 13 agosto 2007]. Disponible en <http://www.w3.org/TR/xpath20/>.
  23. Berners-Lee, T., (ed.) (2006) "Notation3 (N3): A readable language for data on the Web." En línea, [consultado 13 enero 2008]. Disponible en <http://www.w3.org/DesignIssues/Notation3.html>.
  24. Berners-Lee, T. (2000) "Semantic Web - XML2000." En línea, [consultado 13 diciembre 2003]. Disponible en <http://www.w3.org/2000/Talks/1206-xml2k-tbl/>.
  25. Berners-Lee, T. (1994) "W3 future directions". En: 1st World Wide Web Conference. En línea, [consultado 23 noviembre 2007]. Disponible en <http://www.w3.org/Talks/WWW94Tim/>.
  26. Berners-Lee, T.; Fielding, R. y Masinter, L. (2005) "Uniform Resource Identifier (URI): Generic Syntax." En línea, [consultado 6 marzo 2007]. Disponible en <http://gbiv.com/protocols/uri/rfc/rfc3986.html>.
  27. Berners-Lee, T.; Hendler, J. y Lassila, O. (2001) "The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities". The Scientific American (May). En línea, [consultado 16 marzo 2007]. Disponible en <http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>.
-

28. Blanco Fernández, Y. et al. (2004) "AVATAR: An Advanced Multi-Agent Recommender System of Personalized TV Contents by Semantic Reasoning". En: X. Zhong et al. *Web Information System Engineering .Lecture Notes in Computer Science, LNCS*. Vol. 3306, pp. 415-421. Berlin, Heidelberg: Springer Verlag.
  29. Boag, S. et al., (ed.) (2007) "XQuery 1.0: An XML query language." En línea, [consultado 21 diciembre 2007]. Disponible en <http://www.w3.org/TR/xquery/>.
  30. Bollacker, K. D.; Lawrence, S. y Giles, C.L. (2000) "Discovering relevant scientific literature on the web". *IEEE Intelligent Systems* 15(2), pp. 42-47.
  31. Bonissone, P.P. y Decker, K.S. (1986) "Selecting Uncertainty Calculi and Granularity: an experiment in trading-off precision an complexity". En: L.H. Kanal y J.F. Lemmer (eds.) *Uncertainty in artificial intelligence*, pp. 221-247. North-Holland.
  32. Borges, J.L. (1999) "Borges: Collected Fictions" (A. Hurley). New York: Penguin.
  33. Bos, B. et al. (2007) "Cascading Style Sheets, level 2 Revision 1." En línea, [consultado 21 mayo 2008]. Disponible en <http://www.w3.org/TR/CSS21>.
  34. Boutilier, C.; Zemel, R. S. y Marlin, B. (2003) "Active collaborative filtering". En: *Proceedings of the Nineteenth Annual Conference on Uncertainty in Artificial Intelligence*, pp. 98-106.
  35. Boyer, J. M. et al. (2007) "Xforms 1.0 (third edition)." En línea, [consultado 23 enero 2008]. Disponible en <http://www.w3.org/TR/xforms/>.
  36. Bray, T. et al. (eds.) (2006A) "Extensible Markup Language (XML) 1.0 (Forth Edition)." En línea, [consultado 19 febrero 2008]. Disponible en <http://www.w3.org/TR/REC-xml/>.
  37. Bray, T. et al. (eds.) (2006B) "Namespaces in XML." En línea, [consultado 1 septiembre 2008]. Disponible en <http://www.w3.org/TR/REC-xml-names/>.
  38. Brenner, W.; Zarnekow, R. y Witting, H. (1998) "Intelligent Software Agents: Foundations and Applications". Berlin, Heidelberg: Springer Verlag.
  39. Brickley, D. y Guha, R.V. (2004) "RDF Vocabulary Description Language 1.0: RDF Schema." En línea, [consultado 15 agosto 2007]. Disponible en <http://www.w3.org/TR/rdf-schema/>.
  40. Brickley, D. y Miller, L. (2005) "FOAF vocabulary specification." En línea, [consultado 18 junio 2007]. Disponible en <http://www.xmlns.com/foaf/0.1/>.
  41. Bush, V. (1945) "As we may think". *Atlantic Monthly* (176), pp. 101-108.
  42. Buttlar, L. y Tipton, M. (1992) "Library use and staff training in curriculum materials centres". *The journal of academic librarianship* 17 (6), pp. 370-374.
  43. Callan, J. et al. (2003) "Personalisation and Recommender Systems in Digital Libraries". En línea, [consultado 20 abril 2008]. Disponible en [http://www.dli2.nsf.gov/internationalprojects/working\\_group\\_reports/personalisatio](http://www.dli2.nsf.gov/internationalprojects/working_group_reports/personalisatio)
-

*n.html*

44. Cantador, I. y Castells, P. (2006) "Multilayered Semantic Social Network Modeling by Ontology-Based User Profiles Clustering: Application to Collaborative Filtering". En: Managing Knowledge in a World of Networks: Proc. of the 15th International Conference, EKAW 2006, pp. 334-349.
  45. Carmagnola, F. et al. (2006) "The Role of Ontologies in Context-Aware Recommender Systems". En: Proceedings of the 7th International Conference on Mobile Data Management, pp. 80.
  46. Carrizo Sainero, G. (1994) "Manual de fuentes de información". Madrid: CEGAL.
  47. Carzaniga, A.; Rosenblum, D. y Wolf, A.L. (2000) "Interfaces and Algorithms for a Wide-Area Event Notification Service". En línea, [consultado 12 julio 2007]. Disponible en <http://www.cs.colorado.edu/users/carzanig/siena/./papers/cucs-888-99.ps.gz>.
  48. Celma, O.; Ramirez, M. y Herrera, P. (2005) "Foafing the music: A music recommendation system based on RSS feeds and user preferences". En: Proc. of the 6th International Conference on Music Information Retrieval (ISMIR), pp. 464-57.
  49. Christensen, E. et al. (2001) "Web Services Description Languages 1.1." En línea, [consultado 14 enero 2008]. Disponible en <http://www.w3.org/TR/wsdl>.
  50. Clark, K.G.; Feigenbaun, L. y Torres, E., (eds.) (2008) "SPARQL protocol for RDF." En línea, [consultado 2 febrero 2008]. Disponible en <http://www.w3.org/TR/rdf-sparql-protocol/>.
  51. Connolly, D. , (ed.) (2007) "Gleaning Resource Descriptions from Dialects of Languages (GRDDL)." En línea, [consultado 2 marzo 2008]. Disponible en <http://www.w3.org/TR/grddl/>.
  52. Connolly, D. et al. (2001) "DAML+OIL Reference Description." En línea, [consultado 29 febrero 2008]. Disponible en <http://www.w3.org/TR/daml+oil-reference>.
  53. Corcho, O et al. (2003) "Methodologies, tools and languages for building ontologies: Where is their meeting point?" Data & Knowledge Engineering 46, pp. 41-64.
  54. Costello, R. y Sperberg, R. (2001) "XML Schema simpleType Definition of an ISBN" En línea, [consultado 23 febrero 2008]. Disponible en <http://www.xfront.com/isbn.xsd>.
  55. de Bruijn, J. et al. (2005) "D16 The WSMML Specification. WSMML Working Draft ." En línea, [consultado 30 marzo 2008]. Disponible en <http://www.wsmo.org/TR/d16/>.
  56. de la Cueva Martín, M. (1999) "Acceso y utilización de tesauros en Internet". Revista Española De Documentación Científica 22 (4), pp. 531-540.
  57. Decker K.; Sycara K. y Williamson M. (1997) "Middle-Agents for the Internet". En: Proc. of the IJCAI-97, pp. 578-84.
  58. Delgado, M. et al. (2001) "Combining linguistic information in a distributed intelligent agent
-

- model for information gathering on the Internet". En: P.P. Wang (ed.) *Computing with Words*, pp. 251-76. John Wiley & Son.
59. Delgado, M. et al. (2002) "A communication model based on the 2-tuple fuzzy linguistic representation for a distributed intelligent agent system on Internet". *Soft Computing* 6, pp. 320-328.
  60. DeRose, S.; Maler, E. y Daniel, R. (2001) "XML Pointer Language (XPointer) Version 1.0." En línea, [consultado 23 febrero 2008]. Disponible en <http://www.w3.org/TR/WD-xptr>.
  61. DeRose, S.; Maler, E. y Orchard, D. (2001) "XML Linking Language. (XLink) version 1.0." En línea, [consultado 30 enero 2008]. Disponible en <http://www.w3.org/TR/xlink/>.
  62. Díaz-Avilés, V. E. (2005) "Semantic Peer-To-Peer recommender Systems". Tesis no publicada. Universidad Albert-Ludwigs (Freiburg). En línea, [consultado 12 mayo 2008]. Disponible en <http://citeseer.ist.psu.edu/semantic-peer-to-peer1.pdf>.
  63. Ehrig, M.; Haase, P. y Stojanovic, N. (2004) "Similarity for ontologies - a comprehensive framework". En línea, [consultado 11 marzo 2008]. Disponible en <http://www.aifb.uni-karlsruhe.de/WBS/meh/publications/ehrig05similarity.pdf>
  64. Engels, R.H.P. y Lech, T.C.. (2003) "Generating Ontologies for the Semantic Web: OntoBuilder". En: *Towards the Semantic Web*, pp. 91-115 . John Wiley & Sons, Ltd.
  65. Euzenat, J. (2002) "Research Challenges and Perspectives of the Semantic Web". *Intelligent Systems*, IEEE 17 (5), pp. 86-88.
  66. Fabret F. et al. (2001) "Filtering algorithms and implementation for very fast publish/subscribe systems". En: *Proc. of ACM SIGMOD*, pp. 115-126.
  67. Faensen, D. et al. (2001) "Hermes: A Notification Service for Digital Libraries". En: *Proc. of the Joint ACM/IEEE Conference on Digital Libraries (JCDL '01)*, pp. 373-80.
  68. Fallside, D.C. y Walmsley, P., (eds.) (2004) "XML Schema Part 0: Primer Second Edition." En línea, [consultado 10 septiembre 2005]. Disponible en <http://www.w3.org/TR/xmlschema-0/>.
  69. Fazlollahi, B.; Vahidov, R.M. y Aliev R.A. (2000) "Multi-agent distributed intelligent system based on fuzzy decision making". *International Journal of Intelligent Systems* 15, pp. 849-858.
  70. Fensel, D. et al. (2001) "OIL: an ontology infrastructure for the semantic web". *IEEE Intelligent Systems* 16, pp. 38-45.
  71. Fernandez, M.; Gomez-Perez, A. y Juristo, N. (1997) "METHONTOLOGY: from Ontological Art towards Ontological Engineering". En: *Proceedings of the AAAI97 Spring Symposium Series on Ontological Engineering*, pp. 33-40.
  72. Ferraiolo, J., Fujisawa, J. y Jacson, D., (ed.) (2003) "Scalable Vector Graphics (SVG) 1.1 Specification." En línea, [consultado 31 diciembre 2007]. Disponible en <http://www.w3.org/TR/SVG/>.
-

73. Finin, T. y L. Ding. (2006) "Search Engines for Semantic Web Knowledge". En: Proceedings of XTech 2006: Building Web 2.0. En línea, [consultado 08 junio 2008]. Disponible en [http://ebiquity.umbc.edu/\\_file\\_directory\\_/papers/268.pdf](http://ebiquity.umbc.edu/_file_directory_/papers/268.pdf).
  74. Finin, T. y et al. (2005) "Information retrieval and the Semantic Web". En: Proceedings of the 38th Annual Hawaii International Conference on System Sciences. En línea, [consultado 03 marzo 2008]. Disponible en <http://www.cs.umbc.edu/~finin//papers/cikm02/cikm02.pdf>.
  75. Finin, T. y Joshi, A. (2002) "Agents, Trust, and Information Access on the Semantic Web". SIGMOD Record 31 (4), pp. 30-35.
  76. Foltz, P.W. y Dumais, S.T. (1992) "Personalized information delivery: an analysis of information filtering methods". Communications of the ACM 35 (12), pp. 51-60.
  77. Foskett, D. (1997) "Thesaurus". En: K. Spark-Jones y P. Willet (eds.) Readings in Information Retrieval, pp. 111-34. Morgan Kaufmann Publishers.
  78. Miller, G. A. (1956) "The magical number seven or minus two: some limits on our capacity of processing information". Psychological Rev. 63, pp. 81-97.
  79. García Jiménez, A. (2004) "Instrumentos de representación del conocimiento: tesauros versus ontologías". Anales de Documentación 7, pp. 79-95.
  80. García, E. y García, L.A. (2001) "La biblioteca digital". Madrid: Arco Libros.
  81. Geisler G.; McArthur, D. y Giersch, S. (2001) "Developing recommendation services for a digital library with uncertain and changing data". En: Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries, pp. 199-200
  82. Genesereth, M. y Fikes, R. (1992) "Knowledge Interchange Format, version 3.0, reference manual" Tech. Rep. KSL 92-86, En línea, [consultado 01 mayo 2008]. Disponible en <http://www.agent.ai/doc/upload/200302/gene92.pdf>.
  83. Geroimenko, V. y Chen, C. (2003) "Visualizing the Semantic Web: XML-based Internet and Information Visualization". Berlin: Springer Verlag.
  84. Giasson, F. (2007) "Zitgist Search Query Interface: A new search engine paradigm." En línea, [consultado 5 diciembre 2008]. Disponible en <http://fgiasson.com/blog/index.php/2007/03/26/zitgist-search-query-interface-a-new-search-engine-paradigm/>.
  85. Gilchrist, A. (2003) "Thesauri, taxonomies and ontologies-an ethimological note". Journal of Documentation 59 (1), pp. 7-18.
  86. Gnasa, M. et al. (2005) "Personalized Peer Filtering for a Dynamic Information Push". En: Foundations of Intelligent Systems, Vol. 3488/2005, pp. 650-659.
  87. Golbeck, J. (2005) "Semantic Web Interaction through Trust Network Recommender Systems". En: Proc. of the ISWC05 Workshop on End User Semantic Web Interaction. En línea, [consultado 15 agosto 2007]. Disponible en
-

[http://www.ifi.uzh.ch/ddis/fileadmin/events/iswc2005ws/CameraReady/Golbeck\\_TrustInteraction\\_017.pdf](http://www.ifi.uzh.ch/ddis/fileadmin/events/iswc2005ws/CameraReady/Golbeck_TrustInteraction_017.pdf).

88. Golbeck, J.; Parsia B. y Hendler J. (2003) "Trust networks on the semantic web". En: Proceedings of Cooperative Intelligent Agents 2003. En línea, [consultado 23 febrero 2008]. Disponible en <http://www.mindswap.org/papers/CIA03.pdf>.
  89. Gonçalves, M. A. et al. (2004) "Streams, structures, spaces, scenarios, societies (5s): a formal model for digital libraries". ACM Transactions on Information Systems 22 (2), pp. 270-312.
  90. Grant, J. y Becket, D. (2004) "RDF test cases." En línea, [consultado 7 julio 2007]. Disponible en <http://www.w3.org/TR/rdf-testcases/>.
  91. Gruber, T.R. (2005) "Ontology of Folksonomy: A Mash-up of Apples and Oranges." En línea, [consultado 13 junio 2007]. Disponible en <http://tomgruber.org/writing/ontology-of-folksonomy.htm>.
  92. Gruber, T.R. (1995) "Toward principles for the design of ontologies used for knowledge sharing". International Journal of Human-Computer Studies 43 (5-6), pp. 907-928.
  93. Guarino, N. (1998) "Formal ontology and information systems". En: N. Guarino (ed.) Formal Ontology in Information Systems. Proceedings of FOIS'98, pp.3-17 Amsterdam: IOS Press.
  94. Guarino, N. y Welty, C.A. (2004) "An overview to OntoClean". En: S. Staab y R. Studer (eds.) Handbook on ontologies, pp. 151-72. Berlin: Springer Verlag.
  95. Guha, R.; McCool, R. y Miller, E. (2003) "Semantic search". En: 12<sup>th</sup> Int. World Wide Web Conference 2003, pp. 700-709.
  96. Guha, R. et al. (2004) "Propagation of Trust and Distrust". En: International World Wide Web Conference 04, pp. 403-412.
  97. Halpin, H. y Davis, I. (2007) "GRDDL Primer." En línea, [consultado 6 diciembre 2008]. Disponible en <http://www.w3.org/TR/grddl-primer/>.
  98. Hau, J.; Lee, W. y Darlington, J. (2005) "A similarity measure for Semantic Web Services". En: *Proc. of the World Wide Web Conference 2005*. En línea, [consultado 11 abril 2008]. Disponible en <http://www.ai.sri.com/WSS2005/final-versions/WSS2005-Hau-Final.pdf>
  99. Hendler, J. (2001) "Agents and the Semantic Web". IEEE Intelligent Systems (March-April), pp. 30-37.
  100. Hendler, J. (2007) "The dark side of the Semantic Web". IEEE Intelligent Systems (January-February), pp. 2-4.
  101. Hendler, J. (1999) "Is there an intelligent agent in your future?" En línea, [consultado 20 febrero 2005]. Disponible en <http://www.nature.com/nature/webmatters/agents/agents.html>.
-



102. Herman, I. (2007) "Semantic Web Use Cases and Case Studies". En línea, [consultado 12 junio 2008]. Disponible en <http://www.w3.org/2001/sw/>.
  103. Herrera, F. y Herrera-Viedma, E. (1997) "Aggregation Operators for Linguistic Weighted Information". *IEEE Transactions on Systems, Man and Cybernetics* 27, pp. 646-56.
  104. Herrera, F.; Herrera-Viedma, E. y Verdegay, J.L. (1996) "Direct Approach Processes in Group Decision Making using Linguistic OWA operators". *Fuzzy Sets and Systems* 79, pp. 175-190.
  105. Herrera, F. y Martínez, L. (2000) "A 2-tuple fuzzy linguistic representation model for computing with words". *IEEE Transactions on Fuzzy Systems* 8 (6), pp. 746-52.
  106. Herrera, F. et al. (2003) "A Methodology for Generating the Semantics of Unbalanced Linguistic Term Sets". En: 9th International Conference on Fuzzy Theory and Technology, pp. 151-154.
  107. Herrera, F.; Herrera-Viedma E. y Martínez, L. (2000) "A Fusion Approach for Managing Multi-Granularity Linguistic Term Sets in Decision Making". *Fuzzy Sets and Systems* 114, pp. 43-58.
  108. Herrera-Viedma, E. (2001) "Modelling the Retrieval Process of an Information Retrieval System Using an Ordinal Fuzzy Linguistic Approach". *JASIST* 52(6), pp. 460-475.
  109. Herrera-Viedma, E. y Peis, E. (2003) "Evaluating the informative quality of documents in SGML format using fuzzy linguistic techniques based on computing with words". *Information Processing & Management* 39 (2), pp. 195-213.
  110. Herrera-Viedma, E. et al. (2007) "Improvement of Web-based service Information Systems using Fuzzy linguistic techniques and Semantic Web technologies". En: *E-Service intelligence methodologies, technologies and applications*. Vol. 37, pp. 647-666. Berlin: Springer Verlag.
  111. Herrera-Viedma, E. et al. (2005) "A Consensus Support System Model for Group Decision-making Problems with Multigranular Linguistic Preference Relations". *IEEE Trans. on Fuzzy Systems* 13 (5), pp. 644-58.
  112. Hodge, G. (2000) "Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files". Council on Library and Information Resources.
  113. Horrocks, I.; Hendler, J. (2002) "OntoEdit: Collaborative Ontology Engineering for the Semantic Web" En: *Proc. ISWC 2002, LNCS 2342*, pp. 221-235, Berlin: Springer Verlag
  114. Horrocks, I. et al. (2004) "SWRL: A Semantic Web Rule Language Combining OWL and RuleML." En línea, [consultado 17 marzo 2008]. Disponible en <http://www.w3.org/Submission/SWRL/>.
  115. Huang, Z. et al. (2002) "A graph-based recommender system for digital library". En: *Proceedings of the Second ACM/IEEE-CS Joint Conference on Digital Libraries*, pp. 65-73. ACM Press.
-

116. Jaccard, P. (1912) "The distribution of flora in the Alpine zone". *The new phytologist* 11 (2), pp. 37-50.
  117. Jäschke, R. et al. (2007) "Tag Recommendations in Folksonomies". En: *Knowledge Discovery in Databases: Proc. of PKDD07*, pp. 506-514.
  118. Jennings, N.; Sycara, K. y Wooldridge, M. (1998) "A roadmap of agent research and development". *Autonomous Agents and Multi-Agents Systems* 1, pp. 7-38.
  119. Jung, K. et al. (2005) "RDF Triple Processing Methodology for the Recommendation System Using Personal Information". En: *Proc. Of the Int. Conference on Next Generation Web Services Practices*, pp. 241-247.
  120. Kant, K.; Iyer, R. y Mohapatra, P. (2000) "Architectural impact of secure socket layer on Internet servers". En: *Proc. of the Computer Design International Conference 00*, pp. 7-14.
  121. Khare, R. y Tantek C. (2006) "Microformats: a pragmatic path to the semantic web". En: *Proc. of the 15th international conference on World Wide Web*, pp. 865-66.
  122. Khosravi Farsani, H. y Nematbakhsh, M.A. (2006) "Modeling Electronic Product Catalog using OWL Ontology Language". En: *Proceedings of IADIS Conference*.
  123. Kim, S. y Kwon, J. (2007) "Effective Context-aware Recommendation on the Semantic Web". *International Journal of Computer Science and Network Security* 7(8), pp. 154-59.
  124. Kobayashi, I. y Saito, M. (2006) "A study on information recommendation system that provides information related to user's inquiry for information retrieval". En: *Proc. of the 2006 IEEE/WIC/ACM international conference on Web Intelligence and Intelligent Agent Technology*, pp. 385-388.
  125. Kristensen, J. (1993) "Expanding User Queries Using a Search- Aid Thesaurus". *Information Processing and Management* 29 (6), pp. 733-744.
  126. Kruk, S.R. y Decker, S. (2005) "Semantic Social Collaborative filtering with FOAFRealm". En: *Proceedings of the Semantic Desktop, ISWIC2005*. En línea, [consultado 26 marzo 2008]. Disponible en <http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=E6750BF968DB3E05C007DCC560412204?doi=10.1.1.101.1552&rep=rep1&type=pdf>.
  127. Kuflik, T. y Shoval, P. (2000) "Generation of user profiles for information filtering-research agenda". En: *Proceedings of the 23<sup>rd</sup> Annual Int. ACM SIGIR Conference on Research and Development Information Retrieval*, pp. 313-315.
  128. Kuokka, D. y Harada, L. (1995) "Matchmaking for Information Agents". En: *Proc. Of IJCAI'95*, pp. 672-678.
  129. Laliwala, Z.; Sorathia, V. y Chaudhary, S. (2006) "Semantic and Rule Based Event-driven Services-Oriented Agricultural Recommendation System". En: *Proceedings of the 26th IEEE International Conference on Distributed Computing Systems Workshops (ICDCSW'06)*, pp. 24-30.
-

130. Lawrence, S. y Giles, C.L. (1999) "Searching the Web: General and scientific information access". IEEE Communication Magazine 1, pp.116-122.
  131. Levenshtein, V.I. (1966) "Binary codes capable of correcting deletions, insertions, and reversals". *Soviet Physics Doklady* 10 (8), pp. 707-710.
  132. Libby, D. (1999) "RSS 0.9 (Rich Site Summary)". En línea, [consultado 15 mayo 2008]. Disponible en <http://www.purplepages.ie/RSS/netscape/rss0.90.html>.
  133. Lieberman, H. (1999) "Personal assistants for the Web: a MIT perspective". En: Intelligent Information Agents. (ed.) M. Klush, pp. 279-292. Berlin: Springer Verlag.
  134. Loizou, A. y Dasmahapatra, S. (2006) "Recommender Systems for the Semantic Web". En: Proceedings of the ECAI 2006. Recommender Systems Workshop, pp. 76-81.
  135. López Alonso, M.A. (2001) "Integración de teorías para la representación y recuperación del conocimiento. La representación y organización del conocimiento: metodologías, modelos y aplicaciones". En: Actas del V Congreso ISKO-España, pp. 53-67.
  136. Maedche, A. y Staab, S. (2002) "Measuring Similarity between Ontologies". En: *Proc. of the European Conference on Knowledge Acquisition and Management EKAW-2002*. En línea, [consultado 27 junio 2007]. Disponible en <http://www.aifb.uni-karlsruhe.de/~sst/Research/Publications/ekaw2002-compare.pdf>.
  137. Maedche, A. y Volz, R. (2001) "The Ontology Extraction Maintenance Framework Text-To-Onto". Workshop on Integrating Data Mining and Knowledge Management, pp. 177-191
  138. Maes, P. (1994) "Agents that reduce the work and information overload". Communications of the ACM 37 (7), pp. 30-40.
  139. Maes, P. (1995) "Modeling adaptive autonomous agents". En: Artificial life: an overview. Langtonm C.G., pp. 135-62. MIT Press.
  140. Maltz, D. y Ehrlich, K. (1995) "Pointing the way: active collaborative filtering". En: Proceedings of the SIGCHI conference on Human factors in computing systems. ACM Press/Addison-Wesley Publishing, pp. 202-209.
  141. Marchionini, G. (2000) "Research and Development in Digital Libraries." En línea, [consultado 6 agosto 2007]. Disponible en [http://ils.unc.edu/~march/digital\\_library\\_R\\_and\\_D.html](http://ils.unc.edu/~march/digital_library_R_and_D.html).
  142. Martin, D. et al. (2004) "OWL-S: Semantic Markup for Web Services." En línea, [consultado 3 julio 2008]. Disponible en <http://www.w3.org/Submission/OWL-S/>.
  143. Massa, P. y Avesani, P. (2004) "Trust-Aware Collaborative Filtering for Recommender Systems". En: On the Move to Meaningful Internet Systems 2004: CoopIS, DOA, and ODBASE. pp. 3-17. Berlin, Heidelberg: Springer Verlag.
  144. Matthew, R.; Agrawal, R. y Domingos, P. (2003) "Trust Management for the Semantic Web". En: Proceedings of the Second International Semantic Web Conference. En línea,
-

- [consultado 29 mayo 2008]. Disponible en <http://www.cs.washington.edu/homes/pedrod/papers/iswc03.pdf>.
145. McBride, B. (2004) "The Resource Description Framework (RDF) and its vocabulary description Language RDFS". En: S. Staab, y R. Studer (eds.) Handbook on ontologies, pp. 51-66. Berlin: Springer Verlag.
  146. McCray, A.T.; Srinivasan, S. y Browne, A.G. (1994) "Lexical methods for managing variation in biomedical terminologies". En: Proceedings of the 18th Annual Symposium on Computer Applications in Medical Care, pp. 235-239.
  147. McGlashan, S. et al., (ed.) (2004) "Voice Extensible Markup Language (VoiceXML). Version 2.0." En línea, [consultado 12 diciembre 2008]. Disponible en <http://www.w3.org/TR/voicexml20/>.
  148. McGuinness, D.L. (2000) "The Chimaera Ontology Environment". En: Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence, pp. 1123-1124. AAAI Press /The MIT Press.
  149. McGuinness, D.L. y van Harmelen, F., (eds.) (2004) "OWL Web Ontology Language Overview." En línea, [consultado 16 febrero 2005]. Disponible en <http://www.w3.org/TR/2004/REC-owl-features-20040210/>.
  150. McIlraith, S.; Son, T.C. y Zeng, H. (2001) "Semantic Web Services". IEEE Intelligent Systems (March/April), pp. 46-53.
  151. Middleton, S. E. et al. (2002) "Exploiting synergy between ontologies and recommender systems". En: Proc. of the Semantic Web Workshop, pp. 41-50.
  152. Mika, P. (2005) "Ontologies are us: a unified model of social networks and semantics." En línea, [consultado 28 junio 2007]. Disponible en <http://www.cs.vu.nl/~pmika/research/papers/ISWC-folksonomy.pdf>.
  153. Miles, A. y Brickley, D., (eds.) (2008) "SKOS Core Guide." En línea, [consultado 2 julio 2008]. Disponible en <http://www.w3.org/TR/swbp-skos-core-guide/>.
  154. Morales-del-Castillo, J.M.; Herrera-Viedma, E. y Peis, E. (2007) "Modelo semántico-difuso de un sistema de recomendaciones de información para bibliotecas digitales universitarias". II Congreso Español de Informática (CEDI 2007). II Simposio sobre Lógica Fuzzy and Soft Computing (LFSC 2007), pp. 73-80.
  155. Moscoso, P. (2003) "La nueva misión de las bibliotecas universitarias ante el Espacio Europeo de Enseñanza Superior" En línea, [consultado 20 junio 2007]. Disponible en [http://biblioteca.uam.es/documentos/Jornadas\\_REBIUN/4-nueva\\_mision\\_bibliotecas.pdf](http://biblioteca.uam.es/documentos/Jornadas_REBIUN/4-nueva_mision_bibliotecas.pdf).
  156. Moukas, A.; Zacharia, G. y Maes, P. (1999) "Amalthea and Histos: Multiagent systems for WWW sites and representation recommendations". En: Intelligent Information Agents. ed M. Klusch, pp. 293-322. Berlin: Springer Verlag.
-

157. Moya, F. y Peis, E. (2000) "SGML y servicios de información". *El Profesional de la Información* 9 (6), pp. 4-17.
  158. Naaman, M. (2007) "The emerging-semantics web (the semantic web is dead)." En línea, [consultado 19 junio 2007]. Disponible en <http://yahoresearchberkeley.com/blog/2007/05/16/the-emerging-semantics-web-the-semantic-web-is-dead/>.
  159. Neches, R. et al. (1991) "Enabling technology for knowledge sharing". *AI Magazine* 12 (3), pp. 33-56.
  160. Ng, E. (2008) "Louis Monier spreading some FUD on new search directions." En línea, [consultado 13 junio 2008]. Disponible en [http://elliottng.com/elliott-ng/cuills-louis-monier-spreading-some-fud-on-new-search-directions\\_20080304.html](http://elliottng.com/elliott-ng/cuills-louis-monier-spreading-some-fud-on-new-search-directions_20080304.html).
  161. Nottingham, M. (2005) "The Atom Syndication Format ." En línea, [consultado 16 enero 2008]. Disponible en <http://atomenabled.org/developers/syndication/atom-format-spec.php>.
  162. Noy, N.F. y McGuinness, D.L. (2000) "Ontology Development 101: A Guide to Creating Your First Ontology." En línea, [consultado 7 diciembre 2007]. Disponible en [http://protege.stanford.edu/publications/ontology\\_development/ontology101-noy-mcguinness.html](http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html).
  163. Oldakowsky, R. y Byzer, C. (2005) "SemMF: A framework for calculating semantic similarity of objects represented as RDF graphs (Poster of the 4th Semantic Web Conference)." En línea, [consultado 6 julio 2007]. Disponible en [http://www.corporate-semantic-web.de/pub/SemMF\\_ISWC2005.pdf](http://www.corporate-semantic-web.de/pub/SemMF_ISWC2005.pdf).
  164. Orchard, D. (2007) "Guide to versioning XML Languages using new XML Schema 1.1 features." En línea, [consultado 13 junio 2008]. Disponible en <http://www.w3.org/TR/xmlschema-guide2versioning/>.
  165. Otlet, P. (1996) "El tratado de documentación: el libro sobre el libro, teoría y práctica". (Ayuso García, M.D.) Murcia: Universidad de Murcia. (Trabajo original publicado en 1934).
  166. Paice, C.D. (1991) "A thesaural model of information retrieval". *Information Processing & Management* 27 (5), pp. 433-47.
  167. Patel-Schneider, P.F. y Fensel, D. (2002) "Layering the Semantic Web: problems and directions". En: *Proc. of the ISWC*, pp. 16-29.
  168. Pazzani, M.J. (1999) "A framework for collaborative, content-based and demographic filtering". *Artif. Intell. Rev* 13, pp. 393-408.
  169. Peis, E. et al. (2006) "Servicios "semánticos" de difusión selectiva de información (DSI) basados en RSS". 1st International Conference on Multidisciplinary Information Sciences & Technologies (INSCIT 2006), pp. 266-289
  170. Peis, E.; Morales-del-Castillo, J.M. y Delgado-López, J.A. (2008) "Semantic Recommender
-

- Systems.Analysis of the state of the topic". Hipertext.Net 6. En línea, [consultado 12 junio 2008]. Disponible en <http://www.hipertext.net/english/pag1031.htm>.
171. Peis, E; Morales-del-Castillo, J.M. y Herrera-Viedma, E. (2008) "Servicio Semántico de "Difusión Selectiva de Información" (DSI) para bibliotecas digitales". *El Profesional de la Información* (En prensa).
  172. Pemberton, S. et al. (2002) "XHTML 1.0 The extensible Hyper Text Markup Language (Second Edition)." En línea, [consultado 15 septiembre 2007]. Disponible en <http://www.w3.org/TR/xhtml1/>.
  173. Pepper, S. y Moore, G. (2001) "XML Topic Maps (XTM) 1.0 ." En línea, [consultado 19 marzo 2007]. Disponible en <http://www.topicmaps.org/xtm/>.
  174. Petrie, C.J. (1996) "Agent-based engineering: the Web and intelligence". *IEEE Expert* 11 (6), pp. 24-29.
  175. Pollard, R. (1990) "Hypertext presentation of thesauri used in online searching". *Electronic Publishing* 3 (3), pp. 155-72.
  176. Popescul, A. et al. (2001) "Probabilistic models for unified-collaborative and content-based recommendation in sparse-data environments". En: *Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 467-44.
  177. Prud'hommeaux, E. y Seaborne, A., (eds.) (2008) "SPARQL query language for RDF." En línea, [consultado 03/02/2008]. Disponible en <http://www.w3.org/TR/2004/WD-rdf-sparql-query/>.
  178. Pulido, J.R.G. et al. (2006) "Ontology languages for the semantic web: A never completely updated review". *Knowledge-Based Systems* 19 (7), pp. 489-97.
  179. Pérez Agüera, J.R. (2004) "Automatización de tesauros y su utilización en la Web Semántica". *Textos Universitaris De Biblioteconomía i Documentació (BID)* 13. En línea, [consultado 02 septiembre 2007]. Disponible en [http://www2.ub.es/bid/consulta\\_articulos.php?fichero=13perez2.htm](http://www2.ub.es/bid/consulta_articulos.php?fichero=13perez2.htm).
  180. Qin, J. y Paling, S. (2001) "Converting a controlled vocabulary into an ontology: the case of GRM". *Information Research* 6 (2). En línea, [consultado 20 mayo 2007]. Disponible en: <http://informationr.net/ir/6-2/paper94.html>.
  181. Rafter R.; Bradley, K. y Smyth, B. (1999) "Passive Profiling and Collaborative Recommendation". En: *Proceedings of the 10th Irish Conference on Artificial Intelligence and Cognitive Science*. En línea, [consultado 19 marzo 2007]. Disponible en: [www.cs.ucd.ie/staff/bsmyth/CRC/AICS99f.ps](http://www.cs.ucd.ie/staff/bsmyth/CRC/AICS99f.ps)
  182. Reagle, J.M. (2002) "Finding Bacon's Key. Does Google Show How the Semantic Web Could Replace Public Key Infrastructure?" En línea, [consultado 20 diciembre 2007]. Disponible en <http://www.w3.org/2002/03/key-free-trust>.
  183. Resnick, M.L. et al. (2004) "Persuasive Design Through Intelligent Recommendation Systems". En: *Human Factors and Ergonomics Society Annual Meeting Proceeding*. Human
-

Factors and Ergonomics Society, pp. 1503-1507.

184. Resnick, P. (1999) "Semantic Similarity in a Taxonomy: An Information-Based Measure and its Application to Problems of Ambiguity in Natural Language". pp. 95-130.
  185. Resnick, P. et al. (1994) "GroupLens: An open architecture for collaborative filtering of netnews". En: Proceedings of the ACM. Conference on Computer Supported Cooperative Work, pp. 175-186.
  186. Resnick, P. y Varian, H.R. (1997) "Recommender Systems". Special issue of Communications of the ACM 40(3), pp. 56-58.
  187. Roget, P. (1911) "Thesaurus of English words and phrases". Londres: Longmans, Green and Co.
  188. Rorvig, M. (1999) "Images of similarity: a visual exploration of optimal similarity metrics and scaling properties of TREC topic- document sets". Journal of the American Society for Information Science (JASIS), 50 (8), pp. 639-651.
  189. Salton, G. (1971) "The Smart retrieval system experiments". En: Automatic document processing. Salton, G. New York, Englewood Cliffs: PrenticeHall.
  190. Salton, G.; Wong, A. y Yang, C.S. (1975) "A vector space model for automatic indexing". Communications of the Association for Computing Machinery 18(11), pp. 613-620.
  191. Sánchez Tarragó, N. (2007) "Sindicación de contenidos con canales RSS: aplicaciones actuales y tendencias". ACIMED 15 (3). En línea, [consultado 22 mayo 2008]. Disponible en [bvs.sld.cu/revistas/aci/vol15\\_3\\_07/aci03307.htm](http://bvs.sld.cu/revistas/aci/vol15_3_07/aci03307.htm).
  192. Savia, E., Kurki, T. y Jokela, S. (1998) "Metadata Based Matching of Documents and User Profiles". En: Human and Artificial Information Processing, Proc. of the 8th Finnish Artificial Intelligence Conference, pp. 61-70.
  193. Schafer, J.B.; Konstan, J.A. y Riedl, J. (2001) "E-commerce recommendation applications". Data Mining and Knowledge Discovery 5 (1-2), pp. 115-53.
  194. Seaborne, A., (ed.) (2004) "RDQL: A query language for RDF." En línea, [consultado 2 febrero 2005]. Disponible en <http://www.w3.org/Submission/2004/SUBM-RDQL-20040109/>.
  195. Shadbolt, N.; Hall, W. y Berners-Lee, T. (2006) "The Semantic Web revisited". IEEE Intelligent Systems (May-June), pp. 96-101.
  196. Shah, U. et al. (2002) "Information Retrieval on the Semantic". En: Proceedings of the 10<sup>th</sup> International Conference on Information and Knowledge Management, pp. 461-68.
  197. Shapira, B. et al. (1997) "Information filtering: A new two-phase model using stereotypic user profiling". Journal of Intelligent Information Systems 8, pp. 155-65.
  198. Smeaton, A.F. y Callan, J. (2005) "Personalisation and recommender systems in digital libraries". International Journal of Digital Libraries 5, pp. 299-308.
  199. Smith, C. (1997) "Collins Diccionario español-inglés /inglés-español". Barcelona: Grijalbo
-

Mondadori.

200. Soo, V.W. et al. (2003) "Automatic metadata creation: Automated semantic annotation and retrieval based on sharable ontology and case-based learning techniques". En: Proc. of the 3<sup>rd</sup> ACM/IEEE-CS Joint Conference on Digital libraries , pp. 61-72.
  201. Sparck-Jones, K. (2004) "What's new about the Semantic Web? Some questions". En: Proc. of the ACM SIGIR Forum, pp. 18-23.
  202. Stodolsky, D.S. (1990) "Invitational Journals Based Upon Peer Consensus". Roskilde University Centre, Institute of Geography, Socioeconomic Analysis, and Computer Science 29.
  203. Sure, Y.; Stab, S. y Studer, R. (2004) "On-To-Knowledge Methodology (OTKM)". En: R. Studer y S. Staab (eds.) Handbook on ontologies. pp. 117-32. Berlin: Springer Verlag.
  204. Sure, Y. y Studer, R. (2005) "Semantic Web Technologies for Digital Libraries". Library Management 26 (4-5), pp. 190-195.
  205. Sycara, K. et al. (1996) "Distributed intelligent agents". IEEE Expert 11 (6), pp. 36-46.
  206. Szomszor, M. et al. (2007) "Folksonomies, the Semantic Web, and Movie Recommendation". En: Proc. of the ESWC'07, pp. 71-84.
  207. Taylor, R.S. (1986) "Value Added Processes in Information Systems". Norwood: Ablex Publishing Corporation.
  208. Tho, Q.T. y Cao, T.H. (2006) "Automatic Fuzzy Ontology Generation for Semantic Web". IEEE Transactions on Knowledge and Data Engineering 18 (6), pp. 842-856.
  209. Tramullas, J. (2007) "Bibliotecas digitales". En: Actas del VI Seminario de Centros de Documentación Ambiental y Espacios Naturales Protegidos. En línea, [consultado 05 mayo 2008]. Disponible en [http://www.mma.es/porta/cciones/formacion\\_educacion/grupos\\_ceneam/documentacion\\_espacios/pdf/tramullas.pdf](http://www.mma.es/porta/cciones/formacion_educacion/grupos_ceneam/documentacion_espacios/pdf/tramullas.pdf).
  210. Tramullas, J. (2002) "Propuestas de concepto y definición de la biblioteca digital". En: Actas de las II Jornadas de Bibliotecas digitales, pp. 11-20.
  211. van Ossenbruggen, J.; Hardman, L. y Rutledge, L. (2002) "Hypermedia and the Semantic web: A research agenda". Journal of Digital Information 3 (1). En línea, [consultado 01 junio 2008]. Disponible en <http://jodi.tamu.edu/Articles/v03/i01/VanOssenbruggen/>.
  212. Van Rijsbergen, C.J.(1979) "Information Retrieval". London: Butterworths.
  213. Van Slype, G. (1991) "Los lenguajes de indización: Concepción, construcción y utilización en los sistemas documentales". Madrid: Fundación, Germán Sánchez Ruipérez.
  214. Wang, R.Q. y Kong, F.S. (2007) "Semantic-Enhanced Personalized Recommender System". Machine Learning and Cybernetics 7 (19-22), pp. 4069-4074.
  215. White, H.D. (2003) "Pathfinder Networks and Author Cocitation Analysis: A Remapping of
-



- Paradigmatic Information Scientists". *Journal of the American Society for Information Science and Technology* 54 (5), pp. 423-434.
216. Winner, D. (1999) "ScriptingNews 2.0b1". En línea, [consultado 10 Octubre 2007]. Disponible en [http://my.userland.com/stories/storyReader\\$11](http://my.userland.com/stories/storyReader$11).
217. Woerndl, W.; Schueller, C. y Wojtech, R. (2007) "A Hybrid Recommender System for Context-aware Recommendations of Mobile Applications". En: *Proc. of the IEEE 3rd International Workshop on Web Personalisation, Recommender Systems and Intelligent User Interfaces (WPRSIUI'07)*, pp. 871-878.
218. Xin, Z. et al. (2005) "Information Push-Delivery for User-Centered and Personalized Service". En: *Fuzzy Systems and Knowledge Discovery, LNCS Vol. 3613*, pp. 594-602. Berlin: Springer Verlag.
219. Yager, R.R. (1997) "Intelligent agents for World Wide Web advertising decisions". *International Journal of Intelligent Systems* 12, pp. 379-90.
220. Yager, R.R. (2001) "Fusion of multi-agent preference orderings". *Fuzzy Sets and Systems* 112, pp. 1-12.
221. Yager, R.R. (2007) "Centered OWA operators". *Soft Computing* 11 (7), pp. 632-39.
222. Yan, T.W. y García-Molina H. (1999) "The SIFT Information Dissemination System". *ACM Transactions on Database Systems* 24 (4), pp. 199-249.
223. Yu, Z. et al. (2007) "Ontology-Based Semantic Recommendation for Context-Aware E-Learning". En: *Proceedings of the 4th International Conference on Ubiquitous Intelligence and Computing*, pp. 898-907. *Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer Verlag.
224. Zadeh, L.A. (1975) "The concept of a linguistic variable and its applications to approximate reasoning: Part I, Part II, Part III", pp. 199-249 (8), pp. 301-57 (8), pp. 43-80 (9).
225. Zeng, M.L. y Chan, L.M. (2004) "Trends and issues in establishing interoperability among knowledge organization systems". *Journal of the American Society for Information Science and Technology* 55 (5), pp. 377-95.
226. Zermelo, E. (1930) "Über Grenzzahlen und Mengenbereiche". *Fund. Math* 16, pp. 29-47.
227. Zhong, J. et al. (2002) "A conceptual graph matching for semantic search". En: *Proc. of the International Conference on Computer Science (ICCS)*, pp. 92-96
228. Ziegler, C.N. (2004) "Semantic Web Recommender Systems". En: *Current Trends in Database Technology - EDBT 2004 Workshops, LNCS Vol. 3268/2005*, pp. 78-89. Berlin: Springer Verlag.
229. Ziegler, N.; Lausen, G. y Schmidt-Thieme, L. (2004) "Taxonomy driven computation of product recommendations". En: *Proc. of the 30th ACM Conference on Information and Knowledge Management*, pp. 406-415.
-

---

## REFERENCIAS

- [1] PageRank. En línea, [consultado 03 marzo 2008]. Disponible en <http://www.google.com/intl/es/corporate/tech.html>.
- [2] RAE. En línea, [consultado 03 marzo 2008]. Disponible en <http://www.rae.es>.
- [3] WordReference. En línea, [consultado 13 abril 2008]. Disponible en <http://www.wordreference.com/>
- [4] Diccionario Tony Lozano. En línea, [consultado 14 marzo 2008]. Disponible en <http://eubd1.ugr.es/tony/risweb.isa>
- [5] Glosario de términos de Internet. En línea, [consultado 11 enero 2008]. Disponible en <http://www.matisse.net/files/glossary.html>
- [6] Biblioteca universitaria UGR. En línea, [consultado 27 julio 2007]. Disponible en <http://www.ugr.es/~biblio/>
- [7] Web of Knowledge. En línea, [consultado 03 octubre 2007]. Disponible en <http://scientific.thomson.com>
- [8] Science Direct. En línea, [consultado 03 marzo 2008]. Disponible en <http://www.sciencedirect.com/>
- [9] ACM Portal. En línea, [consultado 16 febrero 2008]. Disponible en <http://portal.acm.org>
- [10] Scopus. En línea, [consultado 09 mayo 2008]. Disponible en <http://www.scopus.com/scopus/home.url>
- [11] E-LIS. En línea, [consultado 03 septiembre 2007]. Disponible en <http://eprints.rclis.org/>
- [12] Pajek. En línea, [consultado 29 septiembre 2007]. Disponible en <http://pajek.imfm.si/doku.php>
- [13] Jena. En línea, [consultado 03 marzo 2008]. Disponible en <http://jena.sourceforge.net/>
- [14] Unicode. En línea, [consultado 15 abril 2008]. Disponible en <http://unicode.org/>
- [15] EAD. En línea, [consultado 31 agosto 2008]. Disponible en <http://www.loc.gov/ead/>
- [16] TRIPLE. En línea, [consultado 29 febrero 2008]. Disponible en <http://triple.semanticweb.org/>
- [17] RQL. En línea, [consultado 18 diciembre 2007]. Disponible en <http://139.91.183.30:9090/RDF/RQL/>
- [18] eRQL. En línea, [consultado 11 noviembre 2007]. Disponible en <http://www.dbis.informatik.uni-frankfurt.de/~tolle/RDF/eRQL/main.html>
-

- [19] Xcerpt. En línea, [consultado 28 diciembre 2007]. Disponible en <http://www.xcerpt.org/about/intro/>
- [20] DAML. En línea, [consultado 15 julio 2007]. Disponible en <http://daml.org>
- [21] RuleML. En línea, [consultado 23 marzo 2008]. Disponible en <http://www.ruleml.org/>
- [22] RIF (Rule Interchange Format). En línea, [consultado 11 octubre 2007]. Disponible en [www.w3.org/2005/rules/](http://www.w3.org/2005/rules/)
- [23] Referencias IRI. En línea, [consultado 25 octubre 2007]. Disponible en <http://www.w3.org/International/O-URL-and-ident>
- [24] Biopax. En línea, [consultado 12 diciembre 2007]. Disponible en <http://www.biopax.org>.
- [25] Informes técnicos del W3C. En línea, [consultado 05 marzo 2008]. Disponible en <http://www.w3.org/2002/01/tr-automation/>
- [26] MINDSWAP. En línea, [consultado 17 mayo 2008]. Disponible en <http://www.mindswap.org/>
- [27] SHOE. En línea, [consultado 03 junio 2008]. Disponible en <http://www.cs.umd.edu/projects/plus/SHOE/>
- [28] SchemaWeb. En línea, [consultado 30 marzo 2008]. Disponible en <http://www.schemaweb.info/default.aspx>
- [29] SWOOGLE. En línea, [consultado 28 febrero 2008]. Disponible en <http://swoogle.umbc.edu/>
- [30] iCalendar a RDF. En línea, [consultado 04 marzo 2008]. Disponible en <http://www.w3.org/2002/12/cal/>
- [31] Esquema RDF para calendarios. En línea, [consultado 13 abril 2008]. Disponible en <http://www.w3.org/2002/12/cal/ical.rdf>
- [32] Biodash. En línea, [consultado 09 agosto 2007]. Disponible en <http://www.w3.org/2005/04/swls/BioDash/Demo/>
- [33] SIMILE. En línea, [consultado 13 marzo 2007]. Disponible en <http://simile.mit.edu>
- [34] AGFA on 'Connected Knowledge'. En línea, [consultado 25 diciembre 2007]. Disponible en <http://www.w3.org/2005/04/swls/#agfa>
- [35] Active Semantic Documents. En línea, [consultado 10 junio 2008]. Disponible en <http://lsdis.cs.uga.edu/projects/asdoc/>
- [36] Knowledge Discovery Platform. En línea, [consultado 03 agosto 2007]. Disponible en <http://www.landcglobal.com/index.php>
- [37] Clinical KM. En línea, [consultado 15 enero 2007]. Disponible en <http://www.partners.org/cird/>
-

- 
- [38] POWDER. En línea, [consultado 17 enero 2008]. Disponible en <http://www.w3.org/2007/powder/>
- [39] UMLS. En línea, [consultado 08 enero 2008]. Disponible en <http://www.nlm.nih.gov/research/umls/>
- [40] LOINC. En línea, [consultado 18 agosto 2007]. Disponible en <http://www.regenstrief.org/medinformatics/loinc/>
- [41] Haystack. En línea, [consultado 08 abril 2008]. Disponible en <http://groups.csail.mit.edu/haystack/home.html>
- [42] PLUM. En línea, [consultado 13 enero 2008]. Disponible en <http://plum.csail.mit.edu/>
- [43] JOURKNOW. En línea, [consultado 30 marzo 2008]. Disponible en <http://projects.csail.mit.edu/jourknow/>
- [44] Relo. En línea, [consultado 03 mayo 2008]. Disponible en <http://relo.csail.mit.edu/>
- [45] Eclipse. En línea, [consultado 16 febrero 2008]. Disponible en <http://www.eclipse.org/>
- [46] Piggy Bank. En línea, [consultado 23 julio 2007]. Disponible en [http://simile.mit.edu/wiki/Piggy\\_Bank](http://simile.mit.edu/wiki/Piggy_Bank)
- [47] Solvent. En línea, [consultado 12 marzo 2008]. Disponible en <http://simile.mit.edu/wiki/Solvent>
- [48] Re:Search. En línea, [consultado 07 mayo 2008]. Disponible en <http://simile.mit.edu/exhibit/>
- [49] OntoWeb. En línea, [consultado 27 febrero 2008]. Disponible en <http://www.ontoweb.org>
- [50] Semantic Web Search. En línea, [consultado 03 marzo 2008]. Disponible en <http://www.semanticwebsearch.com/>
- [51] Corese. En línea, [consultado 13 junio 2008]. Disponible en <http://www-sop.inria.fr/teams/edelweiss/wiki/wakka.php?wiki=Corese>
- [52] SWSI. En línea, [consultado 09 junio 2008]. Disponible en <http://www.swsi.org/>
- [53] DBPedia. En línea, [consultado 03 enero 2008]. Disponible en <http://wiki.dbpedia.org/>
- [54] DBin. En línea, [consultado 26 marzo 2008]. Disponible en <http://www.dbin.org/>
- [55] NeOn. En línea, [consultado 17 abril 2008]. Disponible en <http://www.neon-project.org/>
- [56] Knowledge Web. En línea, [consultado 20 mayo 2008]. Disponible en <http://knowledgeweb.semanticweb.org/>
- [57] ESSI. En línea, [consultado 16 marzo 2008]. Disponible en <http://www.essi-cluster.org/>
-

- [58] Tripcom. En línea, [consultado 03 junio 2008]. Disponible en <http://www.essi-cluster.org/projects/tripcom/>
- [59] SUPER. En línea, [consultado 15 septiembre 2007]. Disponible en <http://www.essi-cluster.org/projects/super/>
- [60] DERI. En línea, [consultado 24 junio 2007]. Disponible en <http://www.deri.org/>
- [61] STII. En línea, [consultado 23 junio 2007]. Disponible en <http://www.sti2.org>
- [62] Nepomuk. En línea, [consultado 18 enero 2008]. Disponible en <http://nepomuk.semanticdesktop.org/>
- [63] SIOC. En línea, [consultado 25 agosto 2007]. Disponible en <http://sioc-project.org/>
- [64] MARCONT. En línea, [consultado 03 octubre 2007]. Disponible en <http://www.marcont.org/>
- [65] Calais. En línea, [consultado 25 noviembre 2008]. Disponible en <http://www.opencalais.com/>
- [66] GUID. En línea, [consultado 03 marzo 2008]. Disponible en <http://msdn2.microsoft.com/en-us/library/cc246027.aspx>
- [67] JeromeDL. En línea, [consultado 17 mayo 2008]. Disponible en <http://www.jeromedl.org/>
- [68] Protégé. En línea, [consultado 28 septiembre 2007]. Disponible en <http://protege.stanford.edu/>
- [69] RDFAuthor. En línea, [consultado 22 marzo 2008]. Disponible en <http://rdfweb.org/people/damian/RDFAuthor/>
- [70] IsaViz. En línea, [consultado 13 junio 2008]. Disponible en <http://www.w3.org/2001/11/IsaViz/>
- [71] Ontolingua. En línea, [consultado 09 septiembre 2007]. Disponible en <http://www.ksl.stanford.edu/software/ontolingua/>
- [72] SWOOP. En línea, [consultado 06 julio 2007]. Disponible en <http://code.google.com/p/swoop/>
- [73] SweDE. En línea, [consultado 21 marzo 2008]. Disponible en [http://www.eclipseplugincentral.com/Web\\_Links-index-req-viewlink-cid-251.html](http://www.eclipseplugincentral.com/Web_Links-index-req-viewlink-cid-251.html)
- [74] SMORE. En línea, [consultado 19 febrero 2008]. Disponible en <http://www.mindswap.org/2005/SMORE/>
- [75] Ontostudio. En línea, [consultado 25 marzo 2008]. Disponible en [http://www.ontoprise.de/content/e1171/e1249/index\\_eng.html](http://www.ontoprise.de/content/e1171/e1249/index_eng.html)
- [76] Altova Semantic Works. En línea, [consultado 16 abril 2008]. Disponible en <http://www.altova.com/>
-

- [77] SWAN. En línea, [consultado 10 enero 2008]. Disponible en <http://old.deri.ie/projects/swan/>
- [78] Magpie. En línea, [consultado 14 marzo 2008]. Disponible en <http://kmi.open.ac.uk/projects/magpie/anim.html>
- [79] RDF-Gravity. En línea, [consultado 04 abril 2008]. Disponible en <http://semweb.salzburgresearch.at/apps/rdf-gravity/index.html>
- [80] RDF Room. En línea, [consultado 18 septiembre 2007]. Disponible en <http://www.dfki.uni-kl.de/~grimnes/2006/03/RDFRoom/>
- [81] RDF Validator. En línea, [consultado 03 marzo 2008]. Disponible en <http://www.w3.org/RDF/Validator/>
- [82] Sesame. En línea, [consultado 13 junio 2008]. Disponible en <http://www.openrdf.org/>
- [83] 3Store. En línea, [consultado 21 enero 2008]. Disponible en <http://www.aktors.org/technologies/3store/>
- [84] Kowari. En línea, [consultado 16 marzo 2008]. Disponible en <http://www.kowari.org/>
- [85] Tucana. En línea, [consultado 22 marzo 2008]. Disponible en [lrci.com/pdf/TKS\\_white\\_paper\\_US.pdf](http://lrci.com/pdf/TKS_white_paper_US.pdf)
- [86] YARS. En línea, [consultado 03 enero 2008]. Disponible en <http://sw.deri.org/2004/06/yars/>
- [87] ARC2. En línea, [consultado 17 mayo 2008]. Disponible en <http://arc.semsol.org/home>
- [88] RDFStore. En línea, [consultado 03 marzo 2008]. Disponible en <http://rdfstore.sourceforge.net/>
- [89] Redland. En línea, [consultado 25 abril 2008]. Disponible en <http://librdf.org/>
- [90] RAP. En línea, [consultado 14 abril 2008]. Disponible en <http://www4.wiwiw.fu-berlin.de/bizer/rdfapi/>
- [91] pOWL. En línea, [consultado 06 junio 2008]. Disponible en <http://powl.sourceforge.net/overview.php>
- [92] Perlib. En línea, [consultado 07 abril 2008]. Disponible en <http://www.w3.org/1999/02/26-modules/>
- [93] Pyrple. En línea, [consultado 23 enero 2008]. Disponible en <http://infomesh.net/pyrple/>
- [94] RDFLib. En línea, [consultado 11 junio 2008]. Disponible en <http://rdflib.net/>
- [95] Tabulator. En línea, [consultado 28 julio 2007]. Disponible en <http://www.w3.org/2005/ajar/tab>
- [96] Navegadores Haystack. En línea, [consultado 02 mayo 2008]. Disponible en [http://www.oneclipse.com/plugins/searchplugin/haystack\\_semantic\\_webbrowser/view](http://www.oneclipse.com/plugins/searchplugin/haystack_semantic_webbrowser/view)
-

- [97] Simile Longwell. En línea, [consultado 16 octubre 2007]. Disponible en <http://simile.mit.edu/wiki/Longwell>
- [98] FOAFnaut. En línea, [consultado 03 marzo 2008]. Disponible en <http://www.jibbering.com/foaf/>
- [99] FOAF Explorer. En línea, [consultado 01 enero 2008]. Disponible en <http://xml.mfd-consult.dk/foaf/explorer/>
- [100] PROMT. En línea, [consultado 23 noviembre 2007]. Disponible en <http://protege.stanford.edu/plugins/prompt/prompt.html>
- [101] IBM Web Ontology Manager. En línea, [consultado 16 marzo 2008]. Disponible en <http://www.alphaworks.ibm.com/tech/wom>
- [102] IRS. En línea, [consultado 24 abril 2008]. Disponible en <http://kmi.open.ac.uk/projects/irs/>
- [103] Pellet. En línea, [consultado 03 junio 2008]. Disponible en <http://pellet.owldl.com/>
- [104] FaCT++. En línea, [consultado 10 julio 2007]. Disponible en <http://owl.man.ac.uk/factplusplus/>
- [105] KAON2. En línea, [consultado 23 marzo 2008]. Disponible en <http://kaon2.semanticweb.org>
- [106] CMW. En línea, [consultado 22 marzo 2008]. Disponible en <http://www.w3.org/2000/10/swap/doc/cwm.html>
- [107] RacerPro. En línea, [consultado 10 junio 2008]. Disponible en <http://www.racer-systems.com/products/racerpro/index.phtml>
- [108] SWED. En línea, [consultado 03 enero 2008]. Disponible en <http://www.swed.org.uk>
- [109] Ontoviews. En línea, [consultado 06 enero 2008]. Disponible en <http://www.seco.tkk.fi/projects/semweb/dist.php#ontoviews>
- [110] Portal semántico de Finlandia. En línea, [consultado 09 noviembre 2007]. Disponible en <http://www.suomi.fi/suomifi/suomi/>
- [111] Kango. En línea, [consultado 08 diciembre 2007]. Disponible en <http://www.kango.com/>
- [112] Onthology. En línea, [consultado 26 marzo 2008]. Disponible en <http://www.onthology.org>
- [113] Oyster. En línea, [consultado 03 mayo 2008]. Disponible en <http://oyster.ontoware.org>
- [114] Sindice. En línea, [consultado 14 febrero 2008]. Disponible en <http://sindice.com/query/keyword>
- [115] SWSE. En línea, [consultado 23 julio 2007]. Disponible en <http://www.swse.org/>
-

- [116] Watson. En línea, [consultado 12 diciembre 2007]. Disponible en <http://watson.kmi.open.ac.uk/WatsonWUI/>
- [117] Falcons. En línea, [consultado 16 enero 2008]. Disponible en <http://iws.seu.edu.cn/services/falcons/>
- [118] Ontogator. En línea, [consultado 05 marzo 2008]. Disponible en <http://www.seco.tkk.fi/projects/semweb/dist.php#ontogator>
- [119] Hakia. En línea, [consultado 08 septiembre 2008]. Disponible en <http://www.hakia.com/>
- [120] Semantic Web Search. En línea, [consultado 11 junio 2008]. Disponible en <http://www.semanticwebsearch.com/query/>
- [121] Zitgist. En línea, [consultado 20 abril 2008]. Disponible en <http://zitgist.com/>
- [122] Powerset. En línea, [consultado 14 marzo 2008]. Disponible en <http://www.powerset.com/>
- [123] Microsearch. En línea, [consultado 21 mayo 2008]. Disponible en <http://www.yr-bcn.es/demos/microsearch/>
- [124] Twine. En línea, [consultado 16 febrero 2008]. Disponible en <http://www.twine.com/>
- [125] Gnowsis. En línea, [consultado 03 marzo 2008]. Disponible en <http://www.gnowsis.org/>
- [126] Flickr. En línea, [consultado 06 junio 2008]. Disponible en <http://www.flickr.com/>
- [127] del.icio.us. En línea, [consultado 17 agosto 2007]. Disponible en <http://del.icio.us>
- [128] CiteUlike. En línea, [consultado 29 marzo 2008]. Disponible en <http://www.citeulike.org>
- [129] Machine tag. En línea, [consultado 31 enero 2008]. Disponible en [http://en.wikipedia.org/wiki/Machine\\_tag](http://en.wikipedia.org/wiki/Machine_tag)
- [130] Dublin Core. En línea, [consultado 16 febrero 2008]. Disponible en <http://www.dublincore.org/documents/dces/>
- [131] Creative Commons. En línea, [consultado 12 julio 2007]. Disponible en <http://es.creativecommons.org/>
- [132] eRDF. En línea, [consultado 08 noviembre 2007]. Disponible en <http://research.talis.com/2005/erdf/wiki/Main/RdfInHtml>
- [133] Internet2. En línea, [consultado 16 marzo 2008]. Disponible en <http://www.internet2.edu/>
- [134] Rebiun. 2º Plan estratégico 2007-2010. En línea, [consultado 14 enero 2008]. Disponible en <http://www.rebiun.org/doc/plan.pdf>
- [135] PRISM. En línea, [consultado 26 marzo 2008]. Disponible en <http://www.prismstandard.org/>
-



- [136] MODS. En línea, [consultado 08 enero 2008]. Disponible en <http://www.loc.gov/standards/mods/>
- [137] Channel Definition Format. En línea, [consultado 17 febrero 2008]. Disponible en <http://msdn2.microsoft.com/en-us/library/aa768023.aspx>
- [138] RSS 0.91. En línea, [consultado 09 marzo 2008]. Disponible en <http://backend.userland.com/rss091>
- [139] RSS 0.92. En línea, [consultado 22 noviembre 2007]. Disponible en <http://backend.userland.com/rss092>
- [140] RSS 2.0. En línea, [consultado 15 febrero 2008]. Disponible en <http://cyber.law.harvard.edu/rss/rss.html>
- [141] Ministerio de Cultura. En línea, [consultado 06 marzo 2008]. Disponible en <http://www.mcu.es/RSS/index.html>
- [142] Canal RSS de la Universidad de Granada. En línea, [consultado 08 enero 2008]. Disponible en <http://www.ugr.es/canalRSS.htm>
- [143] Canal RSS de eBay. En línea, [consultado 11 junio 2008]. Disponible en <http://pages.ebay.com/affiliates/tools/rssgenerator/index.html>
- [144] RSS Network. En línea, [consultado 26 febrero 2008]. Disponible en <http://www.rss-network.com/>
- [145] RSS News Feed Factory. En línea, [consultado 28 febrero 2008]. Disponible en <http://www.webdevtips.co.uk/webdevtips/resources/rssdir.php>
- [146] Search4RSS. En línea, [consultado 07 diciembre 2007]. Disponible en <http://www.search4rss.com/index.php>
- [147] Syndic8. En línea, [consultado 10 junio 2008]. Disponible en <http://www.syndic8.com/>
- [148] Autopinger. En línea, [consultado 03 marzo 2008]. Disponible en <http://www.autopinger.com/blog.aspx>
- [149] Pingomatic. En línea, [consultado 22 marzo 2008]. Disponible en <http://pingomatic.com/>
- [150] OPML. En línea, [consultado 12 enero 2008]. Disponible en <http://opml.org/>
- [151] ReFilter. En línea, [consultado 03 junio 2008]. Disponible en <http://re.rephrase.net/filter/>
- [152] Feed Sifter. En línea, [consultado 13 mayo 2008]. Disponible en <http://feedsifter.com/create.php>
- [153] FilterMyRSS. En línea, [consultado 26 febrero 2008]. Disponible en <http://www.filtermyrss.com/>
- [154] Pulverati. En línea, [consultado 20 enero 2008]. Disponible en <http://pulverati.com/>
-

- [155] Feed Collectors. En línea, [consultado 10 marzo 2008]. Disponible en <http://feedcollectors.com/>
- [156] ZAPTXT. En línea, [consultado 16 marzo 2008]. Disponible en <http://zaptxt.com>
- [157] AiderSS. En línea, [consultado 23 enero 2008]. Disponible en <http://www.aiderss.com/>
- [158] Feed Findings. En línea, [consultado 14 abril 2008]. Disponible en <http://www.feedfindings.com/index.html>
- [159] RSSFilter. En línea, [consultado 11 marzo 2008]. Disponible en <http://www.feedforall.com/rssfilter.htm>
- [160] MySyndicaat. En línea, [consultado 22 febrero 2008]. Disponible en [http://www.mysyndicaat.com/mysynd/\\_home](http://www.mysyndicaat.com/mysynd/_home)
- [161] Turtilla. En línea, [consultado 03 marzo 2008]. Disponible en <http://www.turtilla.com/>
- [162] Advanced RSS MlXer. En línea, [consultado 07 mayo 2008]. Disponible en <http://www.advancedrssmixer.com>
- [163] Blasfeed. En línea, [consultado 28 enero 2008]. Disponible en <http://blastfeed.com>
- [164] Yahoo! Pipes. En línea, [consultado 07 mayo 2008]. Disponible en <http://pipes.yahoo.com/pipes/>
- [165] Feedbite. En línea, [consultado 16 marzo 2008]. Disponible en <http://www.feedbite.com/>
- [166] RSS2HTML Scout. En línea, [consultado 25 mayo 2008]. Disponible en <http://www.bytescout.com/rss2htmlscout.html>
- [167] Feed Rinse. En línea, [consultado 03 mayo 2008]. Disponible en <http://feedrinse.com/>
- [168] DOM Initiative Group. En línea, [consultado 14 enero 2008]. Disponible en <http://www.w3.org/DOM/>
- [169] Unified Modeling Language. En línea, [consultado 19 marzo 2008]. Disponible en <http://www.uml.org/>
- [170] Formato de tiempo W3CDTF. En línea, [consultado 23 febrero 2008]. Disponible en <http://www.w3.org/TR/NOTE-datetime>
- [171] ThManager. En línea, [consultado 15 abril 2008]. Disponible en <http://thmanager.sourceforge.net/>
- [172] Tesoro del CINDOC sobre Biblioteconomía y Documentación. En línea, [consultado 03 marzo 2008]. Disponible en [http://thes.cindoc.csic.es/index\\_BIBLIO\\_esp.html](http://thes.cindoc.csic.es/index_BIBLIO_esp.html)
-



---

# ANEXOS



## CÓDIGO DE PERFIL DE USUARIO [EJEMPLO]

---



```
<?xml version="1.0" encoding="iso-8859-1"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:rdfs="http://www.w3.org/
2000/01/rdf-schema#" xmlns:foaf="http://xmlns.com/foaf/0.1/" xmlns:dfss="http://www.ugr.es/
~josemdc/d-fussion/profile#" xml:base="http://www.ugr.es/~josemdc/d-fussion/profile">
<foaf:PersonalProfileDocument rdf:about="">
  <foaf:maker rdf:resource="#persona"/>
  <foaf:primaryTopic rdf:resource="#persona"/>
</foaf:PersonalProfileDocument>
<foaf:Person rdf:ID="user_09235">

<!--Datos personales-->
<foaf:name>Kishimaru Toshi</foaf:name>
<foaf:title>Sr.</foaf:title>
<foaf:givenname>Kisehi</foaf:givenname>
<foaf:family_name>Toshi</foaf:family_name>
<foaf:mbox_sha1sum>af9fa7601d308f46e95596</foaf:mbox_sha1sum>
<foaf:homepage rdf:resource="http://allione.org"/>
<foaf:depiction rdf:resource="allione.jpg"/>
<foaf:phone rdf:resource="tel:555-432-432"/>

<!--Datos de seguridad-->
<dfss:updated>2005-08-17T13:25:42Z</dfss:updated>
<dfss:security_data>
  <dfss:Security_e rdf:nodeID="sec_09235">
    <dfss:login>toshi</dfss:login>
    <dfss:password>jurljurl3</dfss:password>
  </dfss:Security_e>
</dfss:security_data>

<!--Valores de preferencia ponderados-->
  <dfss:preferencias_e>
    <dfss:topics rdf:nodeID="top_09235">

      <dfss:topic>
        <dfss:pref rdf:nodeID="pref_09235-1">
          <rdfs:label>Bibliotecas digitales</rdfs:label>
          <dfss:relev>1.00</dfss:relev>
        </dfss:pref>
      </dfss:topic>
      <dfss:topic>
        <dfss:pref rdf:nodeID="pref_09235-2">
          <rdfs:label>Producción científica</rdfs:label>
          <dfss:relev>0.67</dfss:relev>
        </dfss:pref>
      </dfss:topic>
      <dfss:topic>
        <dfss:pref rdf:nodeID="pref_09235-2">
          <rdfs:label>Tecnologías de la información y la comunicación</rdfs:label>
          <dfss:relev>0.83</dfss:relev>
        </dfss:pref>
      </dfss:topic>
    </dfss:topics>
  </dfss:preferencias_e>

<!--Valoraciones del usuario sobre un recurso determinado-->
  <dfss:valoracion_e>
    <dfss:valoraciones rdf:nodeID="val_09235">

      <dfss:evaluacion>
        <dfss:eval rdf:nodeID="eval_09235-1">
          <dfss:recurso rdf:resource="#doc-00000029"/>
          <dfss:valor>1</dfss:valor>
          <dfss:fecha>2007-12-17T16:15:17Z</dfss:fecha>
        </dfss:eval>
      </dfss:evaluacion>
      <dfss:evaluacion>
        <dfss:eval rdf:nodeID="eval_09235-2">
          <dfss:recurso rdf:resource="#doc-00000528"/>
          <dfss:valor>0.67</dfss:valor>
          <dfss:fecha>2008-01-11T08:41:02Z</dfss:fecha>
        </dfss:eval>
      </dfss:evaluacion>

      <dfss:evaluacion><dfss:eval rdf:nodeID="val_09235-3">
<dfss:recurso rdf:resource="http://www.ugr.es/~josemdc/d-fussion/doc-00005456"/>
<dfss:valor>0.67</dfss:valor>
<dfss:fecha>2008-06-08T19:48:14Z</dfss:fecha>
</dfss:eval>
</dfss:evaluacion>
      <dfss:evaluacion><dfss:eval rdf:nodeID="val_09235-4">
<dfss:recurso rdf:resource="http://www.ugr.es/~josemdc/d-fussion/doc-00010600"/>
```



```
<dfss:valor>0.83</dfss:valor>
<dfss:fecha>2008-06-08T19:47:58Z</dfss:fecha>
</dfss:eval>
</dfss:evaluacion>
<dfss:evaluacion><dfss:eval rdf:nodeID="val_09235-5">
<dfss:recurso rdf:resource="http://www.ugr.es/~josemdc/d-fussion/doc-00008678"/>
<dfss:valor>0.17</dfss:valor>
<dfss:fecha>2008-06-08T19:48:42Z</dfss:fecha>
</dfss:eval>
</dfss:evaluacion>
</dfss:valoraciones>
  </dfss:valoracion_e>

</foaf:Person>
</rdf:RDF>
```

## CÓDIGO DE CANAL RSS [EJEMPLO]

---



```

<?xml version="1.0"?>
<!--<?xml-stylesheet type="text/xsl" href="plantillaRSS.xsl"?-->
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:sy="http://purl.org/rss/1.0/modules/syndication/"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns="http://purl.org/rss/1.0/">
  <channel rdf:about="http://www.ugr.es/~josemdc/d-fussion/canal/">
    <title>Boletín D-Fussion</title>
    <link>http://www.ugr.es/~josemdc/</link>
    <description>Boletín de novedades personalizadas de la biblioteca universitaria</description>
    <dc:creator>José Manuel Morales del Castillo (mailto:josemdc@ugr.es)</dc:creator>
    <dc:contributor>José Andrés Delgado López</dc:contributor>
    <dc:rights>Todos los derechos reservados © 2008</dc:rights>
    <dc:date>2008-05-29T12:00+00:00</dc:date>
    <sy:updatePeriod>hourly</sy:updatePeriod>
    <sy:updateFrequency>2</sy:updateFrequency>
    <sy:updateBase>2008-05-29T12:00+00:00</sy:updateBase>
    <dc:subject xml:lang="es">Biblioteconomía y Documentación</dc:subject>
  </channel>
  <items>
    <rdf:Seq>
      <rdf:li resource="http://www.ugr.es/~josemdc/d-fussion/doc-00000029"/>
      <rdf:li resource="http://www.ugr.es/~josemdc/d-fussion/doc-00000528"/>
      <rdf:li resource="http://www.ugr.es/~josemdc/d-fussion/doc-00001211"/>
      <rdf:li resource="http://www.ugr.es/~josemdc/d-fussion/doc-00002774"/>
      <rdf:li resource="http://www.ugr.es/~josemdc/d-fussion/doc-00003901"/>
      <rdf:li resource="http://www.ugr.es/~josemdc/d-fussion/doc-00004786"/>
      <rdf:li resource="http://www.ugr.es/~josemdc/d-fussion/doc-00005456"/>
      <rdf:li resource="http://www.ugr.es/~josemdc/d-fussion/doc-00006617"/>
      <rdf:li resource="http://www.ugr.es/~josemdc/d-fussion/doc-00007295"/>
      <rdf:li resource="http://www.ugr.es/~josemdc/d-fussion/doc-00008537"/>
      <rdf:li resource="http://www.ugr.es/~josemdc/d-fussion/doc-00008678"/>
      <rdf:li resource="http://www.ugr.es/~josemdc/d-fussion/doc-00009460"/>
      <rdf:li resource="http://www.ugr.es/~josemdc/d-fussion/doc-00010600"/>
      <rdf:li resource="http://www.ugr.es/~josemdc/d-fussion/doc-00010936"/>
      <rdf:li resource="http://www.ugr.es/~josemdc/d-fussion/doc-00011118"/>
    </rdf:Seq>
  </items>
  <image rdf:about="http://www.ugr.es/~josemdc/Recursos/d-fussion_image">
    <title>Logotipo de D-Fussion</title>
    <url>http://www.ugr.es/~josemdc/Recursos/d-fussion.gif</url>
    <link>http://www.ugr.es/~josemdc/</link>
  </image>
  <item rdf:about="http://www.ugr.es/~josemdc/d-fussion/doc-00000029">
    <dc:creator>Barrueco Cruz, José Manuel</dc:creator>
    <title>Reference Linking : un nuevo concepto para facilitar el acceso a la literatura
científica</title>
    <link>http://eprints.rclis.org/archive/00000029/01/citepi.pdf</link>
    <description>La ciencia moderna es una actividad social en la que no existe el trabajo
individualista de una sola persona. Los avances y descubrimientos son fruto del trabajo de grupos
de investigadores que ponen en común sus trabajos por encima de limitaciones geográficas o
temporales. Igual que no existe el investigador individual, tampoco puede existir un documento
científico aislado. Si este tipo de documentos no son más que la presentación en un soporte de los
hallazgos y descubrimientos que los autores han realizado, es lógico que éstos hagan referencia
explícita a cuáles han sido sus puntos de referencia, sus colaboradores, los trabajos que le han
precedido, etc. Dichas menciones se llevan a cabo a través de las citas y las referencias
bibliográficas. De esta forma, cada documento nos habla de otros documentos formando así la inmensa
red que es la literatura científica.</description>
    <dc:date>2002</dc:date>
    <dc:source>El Profesional de la Información, vol. 11, num. 4, pp. 278-282</dc:source>
    <dc:subject xml:lang="es">Citas bibliográficas</dc:subject>
    <dc:subject xml:lang="es">Redes científicas</dc:subject>
  </item>
  <item rdf:about="http://www.ugr.es/~josemdc/d-fussion/doc-00000528">
    <dc:creator>Escudero Sánchez, Manuel</dc:creator>
    <dc:creator>Fernández Cáceres, José Luis</dc:creator>
    <dc:creator>Malgosa Sanahuja, Josemaría</dc:creator>
    <title>Retransmisiones y congresos en Internet</title>
    <link>http://eprints.rclis.org/archive/00000528/02/12_MEscudero_EmissionAudioVideo_good.pdf</
link>
    <description>El presente artículo presenta un breve resumen de la Red Regional de Interconexión
de la Comunidad Autónoma de la Región de Murcia como infraestructura de comunicaciones y servicios
telemáticos básica para el desarrollo de la Sociedad de la Información. Asimismo, presenta las
retransmisiones/difusiones multimedia a través de Internet como uno de los servicios más novedosos
que está red ofrece, tanto en la vertiente de retransmisiones de radio on line como en la de
emisión de eventos puntuales con usuarios registrados (seguimiento de congresos virtuales). El
artículo realiza un repaso de la opción tecnológica utilizada para ofrecer dichos servicios y
detalla los aspectos más importantes de los proyectos donde se han usado los mismos.</description>
    <dc:date>2002</dc:date>
    <dc:source>Contenidos y Aspectos Legales en la Sociedad de la Información (CALSI)
(Conferencia)</dc:source>

```

```
<dc:subject xml:lang="es">Comunicaciones a congresos</dc:subject>
<dc:subject xml:lang="es">Comunidades virtuales</dc:subject>
<dc:subject xml:lang="es">Tecnologías de la información y la comunicación</dc:subject>
</item>

<item rdf:about="http://www.ugr.es/~josemdc/d-fussion/doc-00001211">
  <dc:creator>Torres, Isabel de</dc:creator>
  <title>El feminismo académico en España hoy</title>
  <link>http://eprints.rclis.org/archive/00001211/01/2000-35-63.pdf</link>
  <description>Es un hecho constatable que las temáticas relacionadas con la mujer, o las mujeres, gozan hoy de gran predicamento, como ponen de manifiesto la gran atención y el espacio que les dedican los medios de comunicación en general, autopistas de la información incluidas.</description>
  <dc:date>2000</dc:date>
  <dc:source>Métodos de Información (MEI), vol. 7, num. 35-36, pp. 63-67</dc:source>
  <dc:subject xml:lang="es">Bibliotecas especializadas</dc:subject>
  <dc:subject xml:lang="es">Industria editorial</dc:subject>
  <dc:subject xml:lang="es">Bibliografía</dc:subject>
</item>

<item rdf:about="http://www.ugr.es/~josemdc/d-fussion/doc-00002774">
  <dc:creator>Bereijo Martínez, Antonio</dc:creator>
  <dc:creator>Fuentes Romero, Juan José</dc:creator>
  <title>Los soportes fílmicos, magnéticos y ópticos desde la perspectiva de la conservación de materiales</title>
  <link>http://eprints.rclis.org/archive/00002774/01/a01soportes.pdf</link>
  <description>Las condiciones de conservación de materiales de naturaleza cinematográfica, magnética y óptica, centran la presente investigación. En primer lugar, se analiza el caso de los soportes fílmicos. A tal efecto, tras revisar los hitos principales de su evolución, se expone su composición y se consideran las patologías que sufren, para señalar después las condiciones ambientales para su buena conservación. En segundo término, se examina el tipo de material magnético. Paralelamente al caso anterior, se contempla su evolución, la composición físico-química y las condiciones óptimas de conservación. Finalmente, se aborda el material óptico, viendo su desarrollo histórico, su naturaleza física y las condiciones externas para su adecuada conservación.</description>
  <dc:date>2001</dc:date>
  <dc:source>Anales de Documentación, vol. 4, , pp. 7-37</dc:source>
  <dc:subject xml:lang="es">Conservación de documentos</dc:subject>
  <dc:subject xml:lang="es">Documentos audiovisuales</dc:subject>
  <dc:subject xml:lang="es">Condiciones ambientales</dc:subject>
  <dc:subject xml:lang="es">Soportes ópticos</dc:subject>
  <dc:subject xml:lang="es">Soportes magnéticos</dc:subject>
</item>

<item rdf:about="http://www.ugr.es/~josemdc/d-fussion/doc-00003901">
  <dc:creator>Ortells Montón, Milagros</dc:creator>
  <dc:creator>Sepúlveda Martínez, Gloria</dc:creator>
  <dc:creator>Latorre Zacarés, Ignacio</dc:creator>
  <title>Bibliotecas públicas en comarcas con elevado índice de ruralidad : el ejemplo de la provincia de Valencia</title>
  <link>http://eprints.rclis.org/archive/00003901/01/comunicacion07.pdf</link>
  <description>La comunicación estudia la estructura bibliotecaria en cinco comarcas de la provincia de Valencia con un elevado índice de ruralidad. Se analizan los problemas más frecuentes de las bibliotecas públicas ubicadas en municipios de áreas rurales (malas instalaciones, fondos obsoletos, falta de personal profesional, horarios reducidos, falta de automatización, etc.) y se aportan propuestas de posibles soluciones que contribuyan a dinamizar el servicio bibliotecario en comarcas deprimidas.</description>
  <dc:date>2004</dc:date>
  <dc:source>II Congreso Nacional de Bibliotecas Públicas. La biblioteca pública: compromiso de futuro</dc:source>
  <dc:subject xml:lang="es">Bibliotecas públicas</dc:subject>
  <dc:subject xml:lang="es">Bibliotecas rurales</dc:subject>
  <dc:subject xml:lang="es">Sistemas bibliotecarios</dc:subject>
</item>

<item rdf:about="http://www.ugr.es/~josemdc/d-fussion/doc-00004786">
  <dc:creator>Ubillo Venegas, María Antonieta</dc:creator>
  <dc:creator>Maulén León, Carolina</dc:creator>
  <title>Desarrollo de un portal temático jurídico en Internet : un apoyo a la referencia electrónica</title>
  <link>http://eprints.rclis.org/archive/00004786/01/serie3.pdf</link>
  <description>El bibliotecario como intermediario de información, asume el rol de facilitar y "filtrar" la información pertinente y en el contexto del usuario. En Internet hay muchos webs de contenidos o portales temáticos, en este trabajo se han observado especialmente portales del área jurídica. Se presenta el portal temático como apoyo al servicio de información y se entregan algunas pautas y/o contenidos para organizar en forma exitosa un portal temático, basado en la experiencia de las autoras en el ámbito de la información</description>
  <dc:date>2005</dc:date>
  <dc:source>Serie Bibliotecología y Gestión de Información, num. 3, pp. 1-34</dc:source>
  <dc:subject xml:lang="es">Portales tematicos</dc:subject>
  <dc:subject xml:lang="es">Archivos judiciales</dc:subject>
  <dc:subject xml:lang="es">Recursos electrónicos</dc:subject>
```

```
</item>

<item rdf:about="http://www.ugr.es/~josemdc/d-fussion/doc-00005456">
  <dc:creator>Llueca Fonollosa, Ciro</dc:creator>
  <title>Webs sempre accessibles : les biblioteques nacionals i els dipòsits digitals nacionals =
  Webs siempre accesibles : las bibliotecas nacionales y los depósitos digitales nacionales</title>
  <link>http://eprints.rclis.org/archive/00005456/01/15lluecl.pdf</link>
  <description>Las tecnologías de la información y la comunicación han facilitado que el
  patrimonio cultural, científico y la información en general se presenten en formato digital, así
  como en los formatos analógicos tradicionales. La reacción no se ha hecho esperar, y desde la
  década de los noventa han surgido diversos proyectos destinados a garantizar el acceso permanente a
  la producción digital -su recopilación y almacenaje, el tratamiento, la preservación y la
  difusión-. Se presenta la panorámica mundial de los modelos existentes de depósitos digitales
  nacionales, nombre que reciben estos proyectos impulsados habitualmente por las bibliotecas
  nacionales, con un objetivo común: hacer que las páginas web sean siempre accesibles.</description>
  <dc:date>2005</dc:date>
  <dc:source>BiD : textos universitaris de biblioteconomia i documentació, num. 15</dc:source>
  <dc:subject xml:lang="es">Archivos electrónicos</dc:subject>
  <dc:subject xml:lang="es">Bibliotecas nacionales</dc:subject>
</item>

<item rdf:about="http://www.ugr.es/~josemdc/d-fussion/doc-00006617">
  <dc:creator>Herrero-Solana, Víctor</dc:creator>
  <dc:creator>Morales del Castillo, José Manuel</dc:creator>
  <title>Análisis geopolítico de los mapas de conocimiento</title>
  <link>http://eprints.rclis.org/archive/00006617/01/Analisis-geopolitico-de-los-mapas-de-
  conocimiento.pdf</link>
  <description>En este trabajo hacemos uso de las nuevas técnicas de representación de la
  información, como el análisis de cocitación de sitios web (análisis de cositación) y el
  escalamiento multidimensional (MDS), para analizar de forma gráfica las relaciones que establecen
  entre sí los sitios web de universidades de diferentes países. Pretendemos demostrar que estas
  relaciones siguen un patrón geopolítico en lugar de académico, y que, por tanto, pueden ser
  consideradas cómo auténticos mapas &quot;geopolíticos&quot; de Internet. Este tipo de
  representaciones nos permiten obtener una instantánea de una región geográfica y un momento en el
  tiempo determinados, por lo que podrían ser utilizados como herramienta auxiliar para el análisis
  sociopolítico de la realidad.</description>
  <dc:date>2004</dc:date>
  <dc:source>Arbor: ciencia, pensamiento y cultura, vol. CLXXIX, num. 705, pp. 159-171</dc:source>
  <dc:subject xml:lang="es">Internet</dc:subject>
  <dc:subject xml:lang="es">Webmetría</dc:subject>
  <dc:subject xml:lang="es">Cibermetría</dc:subject>
  <dc:subject xml:lang="es">Representación del conocimiento</dc:subject>
</item>

<item rdf:about="http://www.ugr.es/~josemdc/d-fussion/doc-00007295">
  <dc:creator>Gómez, Nancy D.</dc:creator>
  <title>Constitución de la Biblioteca Digital de la Facultad de Ciencias Exactas y Naturales:
  registro, almacenamiento y puesta a disposición de la producción intelectual de la Facultad</title>
  <link>http://eprints.rclis.org/archive/00007295/01/proyectorepositoriofocn.pdf</link>
  <description>Proyecto donde se diseña el plan y establecen los pasos para la generación de un
  repositorio institucional que contenga la producción intelectual de la Facultad.</description>
  <dc:date>2003</dc:date>
  <dc:source></dc:source>
  <dc:subject xml:lang="es">Producción científica</dc:subject>
  <dc:subject xml:lang="es">Bibliotecas universitarias</dc:subject>
</item>

<item rdf:about="http://www.ugr.es/~josemdc/d-fussion/doc-00008537">
  <dc:creator>Codina, Lluís</dc:creator>
  <title>Evaluación de calidad en sitios web: metodología de proyectos de análisis sectoriales y
  de realización de auditorías (Evaluation of quality in web sites: projects methodology for
  sectorial analyses and audits)</title>
  <link>http://eprints.rclis.org/archive/00008537/01/procedimientos2006.pdf</link>
  <description>The evaluation of digital resources is a discipline of Library and Information
  Sciences that was born in the nineties, when the Web constituted for the first time a credible and
  valuable resource both for academic and professional use. Nevertheless, some valuable resources
  have shared their place in the Web with sites (a) of ridiculous interest, (b) fraudulent, (c)
  plagued of errors or (d) the three things simultaneously. By such reason, as we have indicated, at
  some time from the nineties, emerged the necessity to develop methods that allowed to determine
  what web sites deserved to be listed in directories or data bases of quality digital resources
  offered to users of the academic and professional worlds.</description>
  <dc:date>2007</dc:date>
  <dc:source></dc:source>
  <dc:subject xml:lang="english">Evaluación de la información</dc:subject>
  <dc:subject xml:lang="english">World Wide Web</dc:subject>
  <dc:subject xml:lang="english">Calidad de la información</dc:subject>
</item>

<item rdf:about="http://www.ugr.es/~josemdc/d-fussion/doc-00008678">
  <dc:creator>de León Alvarado, Josefina</dc:creator>
  <dc:creator>Martínez Musiño, Celso</dc:creator>
  <title>Disminuyendo la brecha digital, aportación latinoamericana</title>
```

<link><http://eprints.rclis.org/archive/00008678/01/2006.JdeLeonAlvaradoyCMartinezMusino.brechadigital.pdf></link>  
<description>Se argumenta sobre la brecha digital como la diferencia existente en el uso de las tecnologías de la información y la comunicación (TICs) entre países desarrollados y aquellos llamados emergentes como México y se presentan alternativas para achicar dicha brecha desde la perspectiva latinoamericana y también desde la visión de género.</description>  
<dc:date>2006</dc:date>  
<dc:source>Internacional Know How Conference 2006, weaving the information society: a multicultural and gender perspective</dc:source>  
<dc:subject xml:lang="es">Brecha digital</dc:subject>  
<dc:subject xml:lang="es">Tecnologías de la información y la comunicación</dc:subject>  
<dc:subject xml:lang="es">Sociedad de la información</dc:subject>  
</item>

<item rdf:about="http://www.ugr.es/~josemdc/d-fussion/doc-00009460">  
<dc:creator>Vecchi Pomphile, Analía</dc:creator>  
<title>Tesina "Library Management"; Curso de Post-Grado Management Estratégico Universidad de Belgrano Buenos Aires - Argentina Ciclo 2006</title>  
<link><http://eprints.rclis.org/archive/00009460/01/TesinaAVP.pdf></link>  
<description>Se hace un seguimiento de lo acontecido en el mundo a partir de los avances tecnológicos y específicamente en las unidades de información. Se establecen las comparaciones entre las previsiones efectuadas por especialistas de la información hace casi dos décadas de lo que ocurriría en las bibliotecas del año 2000 y lo que sucede a los centros de información hoy día. Se establecen los nuevos enfoques empresariales a aplicar en el contexto de las estructuras existentes y se determinan los nuevos paradigmas donde se focaliza al cliente como centro de atención y al valor agregado de los productos que ofrecen los servicios de información en las bibliotecas del futuro. Caso Biblioteca y Centro de Documentación Facultad Regional Bahía Blanca.Universidad Tecnológica Nacional.</description>  
<dc:date>2006</dc:date>  
<dc:source></dc:source>  
<dc:subject xml:lang="es">Gestión bibliotecaria</dc:subject>  
<dc:subject xml:lang="es">Industria de la información</dc:subject>  
</item>

<item rdf:about="http://www.ugr.es/~josemdc/d-fussion/doc-00010600">  
<dc:creator>Añover Lopez, Julián Manuel</dc:creator>  
<title>Curricula. Sistema de información sobre la actividad investigadora en un Hospital</title>  
<link><http://eprints.rclis.org/archive/00010600/01/curricula.pdf></link>  
<description>El sistema de información que presentamos gestiona automáticamente la información sobre la actividad investigadora en Ciencias de la Salud desarrollada por el personal contratado en el Hospital 12 de Octubre y el Area XI de Atención Especializada a lo largo del tiempo, con el objetivo de facilitar la construcción de conocimiento sobre la organización. </description>  
<dc:date>2007</dc:date>  
<dc:source></dc:source>  
<dc:subject xml:lang="es">Gestión del conocimiento</dc:subject>  
<dc:subject xml:lang="es">Sistemas de información</dc:subject>  
<dc:subject xml:lang="es">Bases de datos</dc:subject>  
<dc:subject xml:lang="es">Investigadores</dc:subject>  
<dc:subject xml:lang="es">Producción científica</dc:subject>  
</item>

<item rdf:about="http://www.ugr.es/~josemdc/d-fussion/doc-00010936">  
<dc:creator>Campal, F.</dc:creator>  
<dc:creator>Domínguez, R.</dc:creator>  
<title>La formación de usuarios y alfabetización informacional en la BPE de Salamanca "Casa de las conchas"</title>  
<link>[http://eprints.rclis.org/archive/00010936/01/DOSSIER\\_3.pdf](http://eprints.rclis.org/archive/00010936/01/DOSSIER_3.pdf)</link>  
<description>Para la BPE de Salamanca, el servicio de formación de usuarios ha sido siempre prioritario y forma parte del funcionamiento habitual de la misma desde sus inicios. Los primeros programas de formación -en nuestro caso hablamos siempre de adultos, ya que la biblioteca no dispone de Sala Infantil- comenzaron en los años 95/96. Desde entonces se han venido realizando de forma continuada, con los cambios y desarrollos necesarios para adaptarse a las nuevas demandas de la sociedad y a las situaciones concretas de los colectivos asistentes. Aquí se ofrece un recorrido por el pasado y el presente del Servicio de Formación de Usuarios en esta biblioteca.</description>  
<dc:date>2006</dc:date>  
<dc:source>EDUCACIÓN Y BIBLIOTECA, vol. 156, 2006, , pp. 112-116</dc:source>  
<dc:subject xml:lang="es">Formación de usuarios</dc:subject>  
<dc:subject xml:lang="es">Alfabetización en información</dc:subject>  
<dc:subject xml:lang="es">Bibliotecas públicas</dc:subject>  
</item>

<item rdf:about="http://www.ugr.es/~josemdc/d-fussion/doc-00011118">  
<dc:creator>Nardi, Alejandra Marcela</dc:creator>  
<title>El cambio organizacional</title>  
<link>[http://eprints.rclis.org/archive/00011118/01/Microsoft\\_PowerPoint\\_-\\_El\\_cambio.\\_Kit\\_de\\_Transparencias.pdf](http://eprints.rclis.org/archive/00011118/01/Microsoft_PowerPoint_-_El_cambio._Kit_de_Transparencias.pdf)</link>  
<description>En ésta presentación se enuncian los principales conceptos del cambio organizacional: cambios en las personas, resistencia al cambio; cambios no planificados; cambios planificados; cambios impuestos; cambios participativos; cambios negociados. El objetivo del Seminario es inducir a los participantes a aceptar la necesidad de cambiar los procesos y la cultura de la biblioteca.</description>  
<dc:date>1997</dc:date>

```
<dc:source></dc:source>  
<dc:subject xml:lang="es">Bibliotecas</dc:subject>  
<dc:subject xml:lang="es">Gestión bibliotecaria</dc:subject>  
<dc:subject xml:lang="es">Gestión de recursos humanos</dc:subject>  
</item>
```

```
</rdf:RDF>
```





## CÓDIGO DE TESAURO [EJEMPLO]

---



```
<?xml version="1.0" encoding="ISO-8859-1"?>

<rdf:RDF xml:base="http://www.ugr.es/~josemdc/tesauroCINDOC/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">

  <rdfs:Class rdf:ID="tz1"/>
  <rdfs:Class rdf:ID="tz2"/>
  <rdfs:Class rdf:ID="tz3"/>
  <rdfs:Class rdf:ID="tz4"/>
  <rdfs:Class rdf:ID="tz5"/>
  <rdfs:Class rdf:ID="tz6"/>
  <rdfs:Class rdf:ID="tz7"/>
  <rdfs:Class rdf:ID="tz8"/>
  <rdfs:Class rdf:ID="tz9"/>
  <rdfs:Class rdf:ID="tz10"/>
  <rdfs:Class rdf:ID="tz11"/>
  <rdfs:Class rdf:ID="tz12"/>

  <rdfs:Class rdf:ID="t1">
    <rdfs:label xml:lang="es">Accesibilidad</rdfs:label>
    <rdfs:subClassOf rdf:resource="#t1097"/>
    <rdfs:subClassOf rdf:resource="#t1338"/>
    <rdfs:subClassOf rdf:resource="#t1497"/>
    <rdfs:subClassOf rdf:resource="#t1499"/>
  </rdfs:Class>

  <rdfs:Class rdf:ID="t2">
    <rdfs:label xml:lang="es">Acceso a la información</rdfs:label>
    <rdfs:subClassOf rdf:resource="#t334"/>
  </rdfs:Class>

  <rdfs:Class rdf:ID="t3">
    <rdfs:label xml:lang="es">Acceso al documento archivístico</rdfs:label>
    <rdfs:subClassOf rdf:resource="#t388"/>
  </rdfs:Class>

  <rdfs:Class rdf:ID="t4">
    <rdfs:label xml:lang="es">Acceso al documento bibliotecario</rdfs:label>
    <rdfs:subClassOf rdf:resource="#t169"/>
    <rdfs:subClassOf rdf:resource="#t1156"/>
    <rdfs:subClassOf rdf:resource="#t1169"/>
  </rdfs:Class>

  <rdfs:Class rdf:ID="t5">
    <rdfs:label xml:lang="es">Acceso al documento electrónico</rdfs:label>
    <rdfs:subClassOf rdf:resource="#t1097"/>
    <rdfs:subClassOf rdf:resource="#t1338"/>
    <rdfs:subClassOf rdf:resource="#t1497"/>
    <rdfs:subClassOf rdf:resource="#t1499"/>
  </rdfs:Class>

  <rdfs:Class rdf:ID="t6">
    <rdfs:label xml:lang="es">Actas</rdfs:label>
    <rdfs:subClassOf rdf:resource="#t1051"/>
    <rdfs:subClassOf rdf:resource="#t1474"/>
    <rdfs:subClassOf rdf:resource="#t1475"/>
  </rdfs:Class>

  <rdfs:Class rdf:ID="t7">
    <rdfs:label xml:lang="es">Actas de congresos</rdfs:label>
    <rdfs:subClassOf rdf:resource="#t668"/>
  </rdfs:Class>

  <rdfs:Class rdf:ID="t8">
    <rdfs:label xml:lang="es">Actividad museística</rdfs:label>
    <rdfs:subClassOf rdf:resource="#t732"/>
    <rdfs:subClassOf rdf:resource="#t1356"/>
  </rdfs:Class>

  <rdfs:Class rdf:ID="t9">
    <rdfs:label xml:lang="es">Actividad profesional</rdfs:label>
    <rdfs:subClassOf rdf:resource="#t340"/>
  </rdfs:Class>

  <rdfs:Class rdf:ID="t10">
    <rdfs:label xml:lang="es">Actividades culturales archivísticas</rdfs:label>
    <rdfs:subClassOf rdf:resource="#t959"/>
  </rdfs:Class>
</rdf:RDF>
```

```
</rdfs:Class>

<rdfs:Class rdf:ID="t11">
  <rdfs:label xml:lang="es">Actividades culturales bibliotecarias</rdfs:label>
  <rdfs:subClassOf rdf:resource="#t960"/>
</rdfs:Class>

<rdfs:Class rdf:ID="t12">
  <rdfs:label xml:lang="es">Administración electrónica</rdfs:label>
  <rdfs:subClassOf rdf:resource="#t996"/>
</rdfs:Class>

<rdfs:Class rdf:ID="t13">
  <rdfs:label xml:lang="es">Adquisición compartida</rdfs:label>
  <rdfs:subClassOf rdf:resource="#t15"/>
  <rdfs:subClassOf rdf:resource="#t1107"/>
  <rdfs:subClassOf rdf:resource="#t1108"/>
</rdfs:Class>

<rdfs:Class rdf:ID="t14">
  <rdfs:label xml:lang="es">Adquisición de obras</rdfs:label>
  <rdfs:subClassOf rdf:resource="#t493"/>
  <rdfs:subClassOf rdf:resource="#t1104"/>
</rdfs:Class>

<rdfs:Class rdf:ID="t15">
  <rdfs:label xml:lang="es">Adquisiciones</rdfs:label>
  <rdfs:subClassOf rdf:resource="#t339"/>
</rdfs:Class>

<rdfs:Class rdf:ID="t16">
  <rdfs:label xml:lang="es">ADSL</rdfs:label>
  <rdfs:subClassOf rdf:resource="#t82"/>
</rdfs:Class>

<rdfs:Class rdf:ID="t17">
  <rdfs:label xml:lang="es">Agendas electrónicas</rdfs:label>
  <rdfs:subClassOf rdf:resource="#t777"/>
  <rdfs:subClassOf rdf:resource="#t1204"/>
</rdfs:Class>

<rdfs:Class rdf:ID="t18">
  <rdfs:label xml:lang="es">Agentes inteligentes</rdfs:label>
  <rdfs:subClassOf rdf:resource="#t516"/>
</rdfs:Class>

<rdfs:Class rdf:ID="t19">
  <rdfs:label xml:lang="es">Agregadores</rdfs:label>
  <rdfs:subClassOf rdf:resource="#t975"/>
  <rdfs:subClassOf rdf:resource="#t1419"/>
</rdfs:Class>

<rdfs:Class rdf:ID="t20">
  <rdfs:label xml:lang="es">Agrupaciones documentales</rdfs:label>
  <rdfs:subClassOf rdf:resource="#t388"/>
</rdfs:Class>

<rdfs:Class rdf:ID="t21">
  <rdfs:label xml:lang="es">Alfabetización en información</rdfs:label>
  <rdfs:subClassOf rdf:resource="#t474"/>
  <rdfs:subClassOf rdf:resource="#t1250"/>
</rdfs:Class>

<rdfs:Class rdf:ID="t22">
  <rdfs:label xml:lang="es">Almacenamiento de la información</rdfs:label>
  <rdfs:subClassOf rdf:resource="#t854"/>
  <rdfs:subClassOf rdf:resource="#t1392"/>
  <rdfs:subClassOf rdf:resource="#t1178"/>
</rdfs:Class>

<rdfs:Class rdf:ID="t23">
  <rdfs:label xml:lang="es">Almanques</rdfs:label>
  <rdfs:subClassOf rdf:resource="#t484"/>
  <rdfs:subClassOf rdf:resource="#t1240"/>
</rdfs:Class>

<rdfs:Class rdf:ID="t24">
  <rdfs:label xml:lang="es">Altavoces</rdfs:label>
  <rdfs:subClassOf rdf:resource="#t804"/>
</rdfs:Class>

<rdfs:Class rdf:ID="t25">
```

```
<rdfs:label xml:lang="es">Alumnos</rdfs:label>  
<rdfs:subClassOf rdf:resource="#t287"/>  
</rdfs:Class>
```







