

Modelo de servicio semántico de difusión selectiva de información (DSI) para bibliotecas digitales

Por Eduardo Peis, Enrique Herrera-Viedma y José-Manuel Morales-del-Castillo

Resumen: Se presentan las bases teóricas y metodológicas para el desarrollo de un modelo de servicio multiagente de DSI para bibliotecas digitales especializadas que aplica tecnologías de web semántica para gestionar más eficazmente la información, mejorar los procesos de comunicación entre agentes y usuarios, y agilizar el acceso a recursos de interés. Para ello se utilizan canales rss a modo de boletines de novedades a partir de los cuales se generan alertas bibliográficas personalizadas. El servicio dispone de un módulo de gestión de canales rss y otro de push de información. En el primero los documentos son representados en forma de ítems en los canales rss y se les asignan materias semiautomáticamente, equiparando sus palabras clave asociadas con los términos de un tesaurus en formato SKOS Core. En el módulo de push de información se generan las alertas personalizadas de acuerdo a las preferencias definidas en los perfiles de los usuarios.



Eduardo Peis es profesor titular en el Departamento de Biblioteconomía y Documentación de la Universidad de Granada, y es uno de los primeros doctores de la disciplina en la universidad española (desde el año 1997). Está especializado en la investigación de meta-lenguajes como sgml/xml y su aplicación a los servicios de información. Fruto de su tarea docente e investigadora son los numerosos trabajos publicados en revistas científicas del área, centrados principalmente en el estudio de la recuperación de información en la Web aplicando técnicas de soft computing y de web semántica.



Enrique Herrera-Viedma es catedrático de informática del Departamento de Ciencias de la Computación e Inteligencia Artificial de la Universidad de Granada. Sus líneas de investigación incluyen el estudio de la toma de decisiones en grupo, modelos de consenso, modelado lingüístico difuso, agregación de información, recuperación difusa de información, recuperación de información en la Web, bibliotecas digitales, etc. Ha publicado numerosos artículos científicos en revistas y en congresos internacionales y nacionales, y ha editado un libro sobre sus temas de especialización.



José-Manuel Morales-del-Castillo es licenciado en documentación y estudiante de doctorado en el Departamento de Biblioteconomía y Documentación de la Universidad de Granada. Ha publicado diversos trabajos en revistas especializadas en inteligencia computacional, lógica difusa y soft computing. Sus líneas de investigación incluyen la aplicación de tecnologías de web semántica en sistemas de recomendación, bibliotecas digitales y sistemas de e-learning para mejorar la representación, el procesamiento y la recuperación de la información.

Palabras clave: Difusión selectiva de información (DSI), Bibliotecas digitales, Sistemas de recomendación, Web semántica, Rss, Redifusión de contenidos.

Title: A semantic service model of selective dissemination of information (SDI) for digital libraries

Abstract: We present the theoretical and methodological foundations for the development of a multi-agent SDI service model for specialized digital libraries, applying semantic web technologies that permit more efficient information management, improving agent-user communication processes and facilitating accurate access to relevant resources. To do this, rss feeds are used as "current awareness bulletins" to generate personalized bibliographic alerts. The SDI service model has an rss feeds management module and an information push module. In the first module, resources are represented as rss feed items and are also semi-automatically assigned subject terms by matching their associated keywords against the terms of a SKOS Core format thesaurus. In the information push module, bibliographic alerts are customized according to the preferences defined on users' profiles.

Keywords: Selective dissemination of information (SDI), Digital libraries, Recommender systems, Semantic Web, Rss, Syndication.

Peis, Eduardo; Herrera-Viedma, Enrique; Morales-del-Castillo, José-Manuel. "Modelo de servicio semántico de difusión selectiva de información (DSI) para bibliotecas digitales". En: *El profesional de la información*, 2008, septiembre-octubre, v. 17, n. 5, pp. 519-525.

DOI: 10.3145/epi.2008.sep.05

Artículo recibido el 14-12-07

Aceptación definitiva: 11-04-08

1. Introducción

Una de las cuestiones principales que tienen que afrontar tanto los sistemas de información como las bibliotecas o la propia web es la gestión eficaz del gran volumen de documentos que almacenan para poder facilitar a los usuarios un acceso sencillo y ágil. Este problema adquiere una mayor dimensión cuando el usuario presenta un alto grado de especialización y requiere recursos muy específicos, como en el caso de investigadores que buscan documentos de su área científica (**Bollacker; Lawrence; Giles, 2000**).

En el ámbito de las bibliotecas y centros de documentación tradicionalmente se han propuesto diferentes soluciones como por ejemplo la creación de servicios de difusión selectiva de información (DSI). Estos son básicamente sistemas de recomendación basados en contenido (**Popescul, et al., 2001**) en los que de acuerdo con el perfil de los usuarios suscritos se generan periódicamente (o a petición del propio interesado) una serie de alertas en las que se les notifica la existencia de recursos que se adecuan a sus intereses (**Aksoy, et al., 1998**), (**Foltz; Dumais, 1992**).

La introducción de las nuevas tecnologías de la información ha provocado que las bibliotecas se vean en la necesidad de reorientar los servicios que prestan para adaptarlos a este nuevo escenario, en un intento de satisfacer demandas cada vez más específicas (**Marchionini, 2000**). Así por ejemplo servicios como los de DSI han sido objeto de estudio en el área de las bibliotecas digitales (**Faensen, et al., 2001**) o en el entorno de sistemas multiagente (**Decker; Sycara; Williamson, 1997**), (**Kuokka; Harada, 1995**).

En la actualidad muchos servicios DSI se implementan sobre plataformas web con una arquitectura multiagente. Se componen básicamente de una serie de agentes intermediarios que realizan la equiparación de los perfiles de usuario con los documentos y de diferentes agentes de entrada y salida que se encargan de realizar las suscripciones al servicio y presentar las alertas (**Altinel; Franklin, 2000**), (**Carzaniga; Rosenblum; Wolf, 2000**), (**Fabret, et al., 2001**), (**Yan; García-Molina, 1999**). La información generalmente se estructura según un determinado modelo de datos, y los perfiles de usuario se definen utilizando palabras clave que son contrastadas con los descriptores o el texto completo de los documentos.

Estos modelos adolecen no obstante de algunas carencias: 1) los procesos de comunicación entre agentes, y entre agentes y usuarios, se ven dificultados por las diversas formas en que la información está representada; 2) la heterogeneidad en la representación de la información provoca a su vez que ésta no pueda ser reutilizada en otros procesos o por otras aplicaciones;

y 3) generalmente no disponen de mecanismos ágiles y sencillos para generar las alertas.

“La base del servicio es la creación de ‘boletines de novedades’ en los que se recogen recursos novedosos adquiridos por la biblioteca”

Para paliar estas deficiencias es posible mejorar y enriquecer la representación de la información mediante la aplicación de tecnologías del proyecto web semántica (**Berners-Lee; Hendler; Lassila, 2001**) que se presenta como una extensión de la Web actual y que pretende convertirse en una plataforma universal para el intercambio de información. En este nuevo modelo la información está dotada de un significado bien definido que permite una mejor colaboración entre humanos y máquinas (**Berners-Lee, 2000**). 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 ellos a nivel semántico (**Berners-Lee, 2000**), (**Hendler, 2001**).

Por todo ello proponemos la aplicación de algunas de las técnicas de web semántica para implementar un nuevo modelo de servicios DSI para bibliotecas digitales especializadas (en este caso concreto, en el área de biblioteconomía y documentación), basándonos en la plataforma multiagente para acceso a la información que definimos en: **Herrera-Viedma; Peis; Morales-del-Castillo (2006)**.

La base del servicio es la creación de “boletines de novedades” en los que se recogen nuevos recursos o aquellos que por algún otro motivo puedan ser de interés para los usuarios. En ellos se hace una descripción básica de los recursos, que incluye descriptores asignados por los administradores del sistema y que servirán para equiparar los ítems con las preferencias almacenadas en los perfiles de los usuarios. En nuestro caso, para definir estos boletines utilizamos rss 1.0 (**Baged-Dov, et al., 2001a**), un vocabulario que utiliza la sintaxis y estructura de datos de rdf (**Becket, 2004**) diseñado específicamente para gestionar listados de hiperenlaces. El sistema se compone de dos módulos funcionales: uno de gestión de canales rss y otro de *push* de alertas bibliográficas personalizadas.

El primero requiere que los recursos del sistema tengan una estructura definida y un cierto grado de estandarización (en este caso se ha optado por utilizar ar-

```

<Area_Interes rdf:ID="A_Int-pr001">
  <A_Int1_e>
    <a_int1 rdf:ID="a_int1-pr001">
      <area>Biblioteconomía</area>
      <freq>Casi_siempre</freq>
    </a_int1>
  </A_Int1_e>
  <A_Int2_e>
    <a_int2 rdf:ID="a_int2-pr001">
      <area>Archivística</area>
      <freq>Casi_nunca</freq>
    </a_int2>
  </A_Int2_e>

```

Figura 1. Representación de 2 preferencias en un perfil de usuario

títulos científicos), ya que esto facilita en gran medida su representación en los canales rss. La asignación de materias a los ítems se realiza de forma asistida gracias al uso de un tesauro definido en *SKOS Core* (*simple knowledge organisation system*) (Miles; Brickley, 2005), que permite a los agentes sugerir al administrador del sistema las categorías temáticas que mejor describen el contenido del documento de acuerdo con las palabras clave asignadas por su autor. El módulo de *push* por su parte genera las alertas personalizadas según las preferencias definidas en los perfiles de los usuarios.

2. Modelo de servicio semántico de DSI

El servicio de DSI propiamente dicho cuenta con 3 agentes, de interfaz, de canales y de tarea, que se distribuyen en una arquitectura jerárquica de 4 niveles: usuario, interfaz, tarea y recursos.

“La asignación de materias es asistida mediante un tesauro definido en Skos Core que permite a los agentes software sugerir categorías temáticas al administrador del sistema”

La base la forman diferentes técnicas de web semántica que permiten definir cada uno de los elementos principales del sistema: un repositorio de perfiles de usuario; un repositorio de documentos; uno o varios canales rss, que contienen un listado con las novedades bibliográficas; y un tesauro que relaciona términos relevantes dentro del dominio específico sobre el que trabaja la biblioteca. Todos estos componentes están descritos utilizando vocabularios basados en rdf, mientras que su semántica (es decir, sus características y

la relación que mantienen con otros elementos) viene definida en una serie de ontologías web desarrolladas con *OWL* (*ontology web language*) (McGuinness; van Harmelen, 2004).

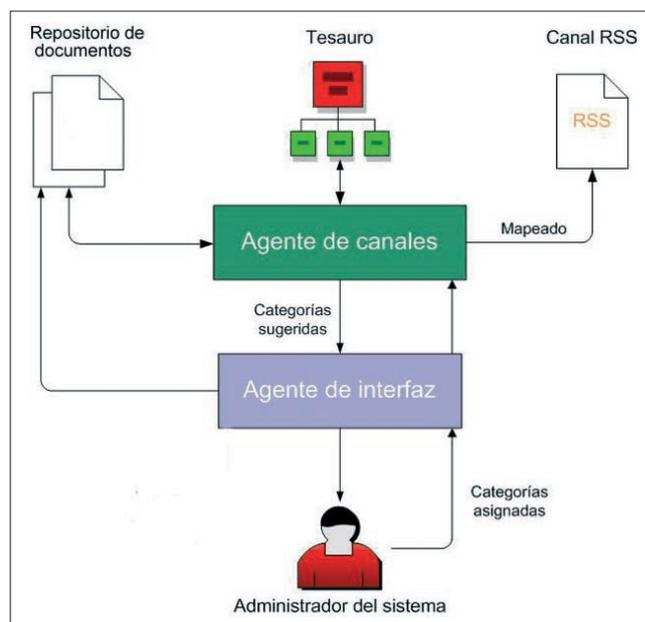


Figura 2. Actualización de canales

La actividad del servicio de DSI se desarrolla en dos módulos funcionales: 1) en el primero se generan y actualizan de forma semiautomática los canales rss gracias al agente de canales, que crea la descripción de los ítems del canal a partir de los metadatos contenidos en los nuevos documentos que ingresan en el sistema y de los términos del tesauro; 2) en el segundo de los módulos, el servicio DSI propiamente dicho, el agente de tarea gestiona la generación de las alertas con acuerdo al perfil de los usuarios.

“El servicio de DSI se compone de dos módulos funcionales: uno de generación semiautomática de canales rss y otro de push de alertas bibliográficas personalizadas”

A continuación se describe cada uno de los elementos principales y los módulos del servicio.

A. Elementos del modelo

A1. Perfiles de usuario. Son representaciones estructuradas que contienen los datos personales, intereses y preferencias de los usuarios con los que los agentes pueden operar para personalizar el servicio de

DSI. Vienen descritos en formato rdf/xml y utilizan el vocabulario FOAF (*friend of a friend*) (Brickley; Miller, 2005), para definir los datos personales (lo que favorece la interoperabilidad del perfil).

En el momento en que una persona se da de alta en el sistema, ha de proporcionar algunos datos personales y sus preferencias con respecto a la materia de las alertas que desea recibir, definidas en el apartado “Área de interés”. Éstas son seleccionadas de entre las 12 categorías principales en las que se divide el tesoro (ver elemento A4). En el perfil de usuario cada una de estas categorías tiene asociada una frecuencia lingüística (etiquetada como $\langle freq \rangle$), que representa la asiduidad con la que es usada en consultas y valoradas por el usuario como satisfactorias (Herrera-Viedma; Peis; Morales-del-Castillo, 2006).

El rango de posibles valores para las frecuencias está definido por un conjunto de 7 etiquetas lingüísticas $S = \{siempre, casi_siempre, mayoría_de_veces, a_veces, pocas_veces, casi_nunca, nunca\}$. Por lo tanto, cuando el módulo de DSI lance una consulta, el sistema usará aquellos valores de preferencia que mayor frecuencia asociada tengan. 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 de éstos debe de apoyarse en algún tipo de *feedback* que permita determinar el comportamiento del individuo en su interacción con el sistema.

En el modelo propuesto por Herrera-Viedma, Peis y Morales-del-Castillo (2006), los usuarios tras realizar una consulta a la biblioteca deben evaluar el resultado obtenido con un grado de satisfacción lingüístico e_j , que nos ofrece una idea aproximada de la “bondad” de los valores de preferencia seleccionados para definir dicha consulta. De esta manera el sistema puede seleccionar aquellos valores de preferencia que han proporcionado respuestas más satisfactorias para un usuario

“Para actualizar los perfiles de usuario se utiliza una función de equiparación que bonifica las frecuencias de los valores de preferencia usados en consultas satisfactorias”

determinado, en lugar de los que han sido seleccionados más frecuentemente. Es decir, lo que se pretende es descubrir las preferencias tácitas de los usuarios. Este grado de satisfacción global e_j se puede definir utilizando un conjunto de 7 etiquetas lingüísticas $S' = \{total, muy_alto, alto, medio, bajo, muy_bajo, nulo\}$.

Para actualizar las frecuencias se utiliza una función de equiparación similar a las aplicadas para modelar pesos de umbral en consultas ponderadas (Herrera-Viedma, 2001), que bonifica las frecuencias de los valores de preferencia usados en consultas satisfactorias y penaliza los usados en consultas no satisfactorias.

A2. Documentos. El conjunto de documentos disponibles en la biblioteca digital a texto completo deben estar codificados en formato xml o html y pueden estar almacenados en un repositorio local o bien estar accesibles de forma remota en la Red. El sistema requiere trabajar con descripciones de estos recursos (que se almacenan en forma de ítems en el canal rss) y en las que entre otros datos se incluye el título, autor/es, resumen, palabras clave y el enlace al documento completo. Estas descripciones se pueden generar manualmente o de forma semiautomática utilizando herramientas de extracción de metadatos y *scripts* de transformación del vocabulario nativo a rss.

A3. Canales rss. En los últimos años gracias a la popularización de los blogs se está imponiendo el uso del vocabulario rss como medio para la redifusión o sindicación de contenidos (hacer accesible a otros usuarios de internet el contenido de un sitio web mediante listas de enlaces denominadas canales o *feeds*).

La estructura de los canales comprende dos áreas básicas: una en la que se describe el canal en sí y otra en la que se listan los recursos o ítems. Los canales pueden ser ficheros xml o rdf/xml dependiendo de la versión rss que se decida utilizar. En este caso utilizamos rss 1.0 (*rdf site summary*), que se basa en el modelo de datos y sintaxis de rdf/xml. A pesar de que 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 más extendido en la Red), rss 1.0 presenta una estructura de datos más compleja que las versiones anteriores, pero ofrece mayores capacidades de interoperabilidad y extensibilidad. Esto se debe al uso de módulos, los cuales definen etiquetas específicas que permiten ampliar el vocabulario rss, sin necesidad de modificar el núcleo de la especificación cada vez que se quieran añadir nuevos elementos descriptivos.

En este modelo se utiliza el módulo DC (Baged-Dov, et al., 2000) para definir la información bibliográfica básica de los ítems utilizando los elementos establecidos por la DCMI (*Dublin core metadata initiative*)¹; el módulo *syndication* (Baged-Dov, et al., 2001b) para facilitar a los agentes la sincronización y la actualización del canal; y el módulo *taxonomy* (Baged-Dov, et al., 2001c) para asignar materias a los ítems.

A4. Tesoro. El modelo de servicio DSI requiere de un tesoro que represente los conceptos más relevantes

del dominio en el que esté especializada la biblioteca digital, para asistir en la tarea de indización de los ítems del canal rss. Aquí se utiliza un tesauro especializado en el área de biblioteconomía y documentación, que segmenta el dominio 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. Estas categorías son el dominio de expresión tanto de las materias de clasificación de los ítems del canal rss, como de la preferencia “Área de interés” en el perfil de los usuarios.

El hecho de utilizar una baja granularidad pretende facilitar al máximo, tanto al usuario como al administrador, las tareas de creación de perfiles y de indización de recursos respectivamente, aunque esto signifique una pérdida de precisión con respecto a otros sistemas con un mayor nivel de detalle.

El tesauro utilizado es una versión del *Tesauro de Biblioteconomía y Documentación del Cindoc*², definida con *Skos core*, un vocabulario rdf específico para migrar tesauros a la web que permite definir las relaciones jerárquicas y asociativas entre los conceptos del dominio, a modo de ontología simple. Existe un sitio web³ de acceso libre donde se puede consultar esta versión *Skos core* del tesauro.

B. Módulos funcionales del sistema

B1. Módulo de gestión de canales. En este módulo se crean y actualizan de manera semiautomática los canales rss de la biblioteca digital (uno o varios canales dependiendo de sus necesidades específicas). Para el proceso de creación el administrador dispone de una interfaz a través de la cual genera la descripción básica del canal: título, breve descripción del contenido, frecuencia de actualización, etc.

A este canal básico se añaden posteriormente las descripciones de los ítems durante el proceso de actualización, las cuales son generadas de forma semiautomática a partir de los documentos en formato xml o html. El proceso de actualización del canal se desarrolla en los siguientes pasos:

– Periódicamente el agente de canales comprueba el repositorio de documentos en busca de aquellos que se han almacenado recientemente y los selecciona para generar su correspondiente representación en el canal rss. Cada ítem del canal requiere básicamente 6 campos de información: título, autor, resumen, enlace al recurso, fecha de creación del ítem y clasificación por materias (figura 3). El agente de tarea intentará extraer del documento el contenido de los tres primeros campos, estableciendo a continuación una correspondencia entre las etiquetas que representan el título, autor y resumen del recurso con las etiquetas rss del ítem del canal. El enlace al recurso y la fecha de creación se generan en el mismo momento de la actualización.

```
<item rdf:about="http://www.bibExample.org/ies/Bib">
  <title>Biblioteconomía aplicada</title>
  <link>http://www.bibExample.org/ises/Bib.pdf</link>
  <description>Resumen del recurso</description>
  <dc:creator>Marco Allesi</dc:creator>
  <dc:date>2005-03-17T13:25:42Z</dc:date>
  <taxo:topic rdf:about="http://www.thesExample.net/CINDOC/169">
  <dc:subject>Biblioteconomía</dc:subject>
  </taxo:topic>
</item>
```

Figura 3. Ejemplo de ítem en el canal rss

Aquellos datos que no hayan podido ser extraídos, que no estén completos o sean incorrectos, deberán ser completados o modificados de forma manual por el propio administrador.

– Posteriormente se procede a la asignación semiautomática de materias de la siguiente forma: el agente de canales equipara el área de palabras clave extraídas del documento con las entradas del tesauro, utilizando para ello el algoritmo *edit-tree* (Levenshtein, 1966) que permite calcular su similitud léxica. Si la equiparación es positiva el agente procede a extraer por cada uno de los términos del tesauro con los que ha habido correspondencia, las respectivas categorías temáticas principales de las que dependen jerárquicamente.

– Estas categorías principales son sugeridas como posibles materias al administrador del sistema que, siguiendo su propio criterio, puede decidir asignarlas o no para representar el contenido del documento en el ítem del canal.

B2. Módulo de *push* de información. Realiza la función de DSI propiamente dicha. Su funcionamiento consiste básicamente en realizar consultas pasivas en lugar del usuario al canal rss de la biblioteca sobre un determinado tema de interés de acuerdo a las preferencias definidas y posteriormente generar una alerta personalizada que es presentada al usuario sin que éste lo

haya requerido expresamente (lo que se conoce como *push* o recepción pasiva de información). La actividad de este módulo engloba los siguientes pasos:

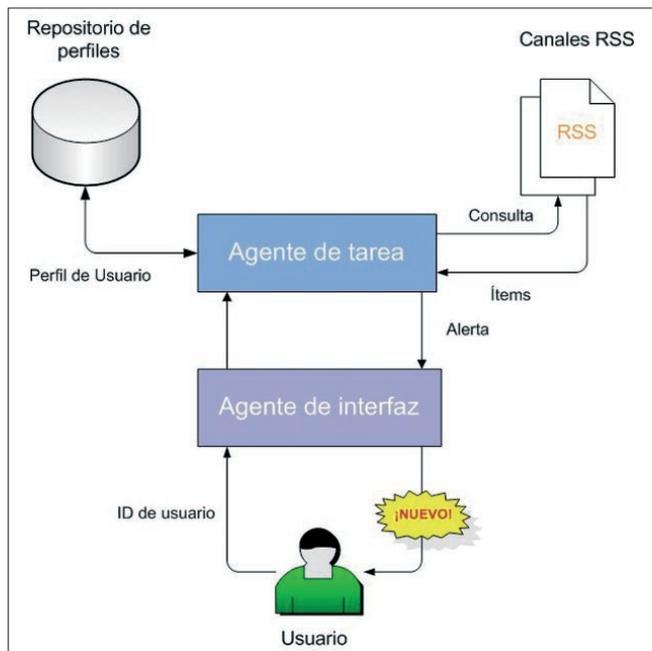


Fig. 4. Módulo de push de información

- El usuario accede al sistema introduciendo su nombre de usuario y clave. Una vez identificado positivamente, el agente de interfaz recupera su perfil.

- A continuación el agente de tarea dispara una consulta pasiva, utilizando *Sparql (Prud'hommeaux; Seaborne, 2006)*, un lenguaje de interrogación que permite realizar búsquedas en grafos rdf, para comparar las preferencias almacenadas en el perfil del usuario activo con los descriptores que caracterizan el contenido de los ítems del canal. El resultado es un listado de ítems que satisfacen la consulta. 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.

- Por último, 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 generado por el agente de tarea, desde el cual el usuario puede acceder a los diferentes recursos. En caso de no hallarse ítems que satisfagan la consulta se notifica al usuario que no existen novedades.

3. Conclusiones

Las bibliotecas se están trasladando a la web, y con ellas los servicios que prestan a sus usuarios, como los servicios de DSI. La aplicación de técnicas de web se-

mántica permite trabajar de una manera más eficiente con la información y el conocimiento, lo que posibilita el desarrollo de servicios DSI mejorados que ofrecen información personalizada más acorde con las necesidades de los usuarios.

“Usando técnicas de web semántica se obtiene un modelo de datos y una sintaxis comunes que garantizan la interoperabilidad y reutilización de los recursos del sistema”

Las técnicas de web semántica permiten tener un modelo de datos y una sintaxis comunes que garantizan la interoperabilidad y reutilización de los recursos del sistema (independientemente de la plataforma en la que se trabaje), facilitando así el establecimiento de redes de colaboración e intercambio con otras bibliotecas digitales y mejorando sensiblemente los procesos de comunicación entre agentes, y entre usuarios y agentes.

En el modelo que aquí presentamos se utilizan algunas de las tecnologías aludidas, como *rdf* para definir documentos y perfiles, *Skos core* para describir el tesoro que ayuda en tareas de indización, *OWL* para desarrollar las ontologías que definen semánticamente los diferentes elementos del modelo y *rss* para crear canales de redifusión de contenidos a modo de boletines de novedades a partir de los cuales generar alertas personalizadas de una manera flexible y sencilla.

La aproximación a la recuperación de ítems de los canales rss se hace cambiando la perspectiva del usuario: de una posición activa en la que el individuo debía definir una serie de palabras clave o navegar entre diferentes categorías temáticas para elegir aquellos canales rss que pueden contener ítems de su interés, se pasa a una posición pasiva en la que la consulta la realiza de forma implícita el sistema utilizando las preferencias del usuario almacenadas en su perfil personal para filtrar los contenidos.

Otra ventaja adicional para la biblioteca digital es la posibilidad de usar los canales rss (desarrollados específicamente para el servicio de DSI) de forma convencional; es decir, hacerlos accesibles a cualquier usuario de internet que se suscriba al canal y disponga de un agregador compatible con rss 1.0, lo cual redundará en una mayor visibilidad de la biblioteca en la Web y una mayor difusión de sus recursos.

Futuros trabajos incluyen el aumento de las capacidades del sistema dotándolo de un mecanismo de

equiparación semántica de ítems y perfiles, así como aumentar el nivel de especificidad para definir preferencias y materias. Otra línea de trabajo consiste en estudiar la adaptación de este tipo de sistemas a las bibliotecas digitales universitarias en su transformación en *centros de recursos para el aprendizaje y la investigación (CRAI)* dentro del *Espacio común europeo de educación e investigación*. En este ámbito ya no sólo sería necesario tener en cuenta las preferencias del usuario sino también sus propias habilidades y nivel de conocimientos.

Agradecimientos

Este trabajo se enmarca dentro del *Proyecto de Excelencia Sainfoweb 0602* de la *Junta de Andalucía* y del proyecto *Fuzzy-Ling (TIN2007-61079)* del *Ministerio de Educación y Ciencia*.

Notas

1. <http://dublincore.org/>
2. http://thes.cindoc.csic.es/index_BIBLIO_esp.html
3. <http://www.ugr.es/~josemdc/tesauroCINDOC/ThesCINDOC.rdf>

Bibliografía

- Aksoy, D.**, et al. "Research in data broadcast and dissemination". En: *Proceedings of the 1st International conference on advanced multimedia content processing*, 1998, pp. 194-207.
- Altinel, M.; Franklin M. J.** "Efficient filtering of xml documents for selective dissemination of information". En: *Proceedings of the 26th VLDB conference*, 2000, pp. 53-64.
- Becket, D.** (ed.). *Rdf/xml syntax specification (revised)*, 2004. Consultado en: 01-02-07. <http://www.w3.org/TR/rdf-syntax-grammar/>
- Beged-Dov, G.**, et al. *Rdf site summary 1.0 modules: Dublin core*, 2000. Consultado en: 31-01-07. <http://web.resource.org/rss/1.0/modules/dc/>
- Beged-Dov, G.**, et al. (b). *Rdf site summary (rss) 1.0*, 2001a. Consultado en: 30-01-07. <http://web.resource.org/rss/1.0/spec>
- Beged-Dov, G.**, et al. (c). *Rdf site summary 1.0 modules: syndication*, 2001b. Consultado en: 02-02-07. <http://web.resource.org/rss/1.0/modules/syndication/>
- Beged-Dov, G.**, et al. (d). *Rdf site summary 1.0 modules: taxonomy*, 2001c. Consultado en: 31-01-07. <http://web.resource.org/rss/1.0/modules/taxonomy/>
- Berners-Lee, T.** *Semantic web road map*, 2000. Consultado en: 29-01-07. <http://www.w3.org/DesignIssues/Semantic.html>
- Berners-Lee, T.; Hendler, J.; Lassila, O.** "The semantic web: a new form of web content that is meaningful to computers will unleash a revolution of new possibilities". En: *Scientific American*, 2001, May.
- Bollacker, K. D.; Lawrence, S.; Giles, C. L.** "Discovering relevant scientific literature on the web". En: *IEEE intelligent systems*, 2000, v. 15, n. 2, pp. 42-47.
- Brickley, D.; Miller, L.** *FOAF vocabulary specification*, 2005. Consultado en: 28-01-07. <http://www.xmlns.com/foaf/0.1/>
- Carzaniga, A.; Rosenblum, D.; Wolf, A. L.** "Interfaces and algorithms for a wide-area event notification service". En: *Proceedings of the 19th ACM PODC*, 2000.
- Cayzer, S.** "Semantic blogging and decentralized knowledge management". En: *Communications of the ACM*, 2004, v. 47, n. 12, pp. 47-52.
- Decker, K.; Sycara, K.; Williamson, M.** "Middle-agents for the internet". En: *Proceedings of Ijcai-97*, 1997, pp. 578-584.
- Fabret, F.; Jacobsen, H. A.; Llibat, F.; Pereira, J.; Ross, K. A.; Shasha, D.** "Filtering algorithms and implementation for very fast publish/subscribe systems". En: *Proceedings of ACM Sigmod*, 2001, 115-126.
- Faensen, D.; Faulstich, L.; Schweppe, H.; Hinze, A.; Steindinger, A.** "Hermes: a notification service for digital libraries". En: *Proceedings of the Joint ACM/IEEE conference on digital libraries (JCDL '01)*, 2001, pp. 373-380.
- Foltz, P. W.; Dumais, S. T.** "Personalized information delivery: an analysis of information filtering methods". En: *Communications of the ACM*, 1992, v. 35, n. 12, pp. 51-60.
- Hendler, J.** "Agents and the semantic web". En: *IEEE intelligent systems*, 2001, March/April, pp. 30-37.
- Herrera-Viedma, E.** "Modelling the retrieval process of an information retrieval system using an ordinal fuzzy linguistic approach". En: *Journal of the American Society for Information Science and Technology (Jasist)*, 2001, v. 52, n. 6, pp. 460-475.
- Herrera-Viedma, E.; Peis, E.; Morales-del-Castillo, J. M.** "A fuzzy linguistic multi-agent model based on semantic web technologies and user profiles". En: **Herrera-Viedma, E.; Pasi, G.; Crestani, F.** *Soft computing in web information retrieval*. Berlin/Heidelberg: Springer, 2006, pp 105-120.
- Kuokka, D.; Harada, L.** "Matchmaking for information agents". En: *Proceedings of Ijcai'95*, 1995, pp. 672-678.
- Levenshtein, V. I.** "Binary codes capable of correcting deletions, insertions, and reversals". En: *Soviet physics doklady*, 1966, v. 10, n. 8, pp. 707-710.
- McGinnes, D. L.; van Harmelen, F.** *OWL web ontology language overview*, 2004. Consultado en: 31-01-07. <http://www.w3.org/TR/owl-features/>
- Marchionini, G.** *Research and development in digital libraries*, 2000. Consultado en: 03-02-07. http://ils.unc.edu/~march/digital_library_R_and_D.html
- Miles, A.; Brickley, D.** *SKOS core guide*, 2005. Consultado en: 29-01-07. <http://www.w3.org/TR/2005/WD-swbp-skos-core-guide-20051102/>
- Popescul, A.; Ungar, L. H.; Pennock, D. M.; Lawrence, S.** "Probabilistic models for unified-collaborative and content-based recommendation in sparse-data environments". En: *Proceedings of the Seventeenth conference on uncertainty in artificial intelligence (UAI)*, 2001, pp. 437-444.
- Prud'hommeaux, E.; Seaborne, A.** (eds.). *Sparql query language for rdf*, 2006. Consultado en: 02-02-07. <http://www.w3.org/TR/rdf-sparql-query/>
- Yan, T. W.; García-Molina, H.** "The SIFT information dissemination system". En: *ACM transactions on database systems*, 1999, v. 24, n. 4, pp. 529-565.

Eduardo Peis, Enrique Herrera-Viedma*, José-Manuel Morales-del-Castillo, Dpto. de Biblioteconomía y Documentación, Univ. de Granada, Granada.

*Dpto. de Ciencias de la Computación e Inteligencia Artificial, Univ. de Granada, Granada.

epais@ugr.es

viedma@decsai.ugr.es

josemdc@ugr.es