

SABIOS: una aplicación de la Web semántica para la gestión de documentos digitales*

Jaime A. Guzmán Luna**
Durley Torres Pardo***
Demetrio A. Ovalle****

Resumen

La Web semántica, pretende convertir la información en conocimiento. Para ello, se basa fundamentalmente en el marcado semántico y descriptivo no sólo de los documentos, sino también de los datos, por medio de información estructurada y legible automáticamente, que proporciona a las máquinas una mayor capacidad para gestionar y recuperar esos datos. SABIOS, se propone precisamente, la introducción de las tecnologías emergentes basadas en semántica, combinadas con sistemas Multiagentes y técnicas de recuperación información para permitir mejorar los procesos de inserción, catalogación y recuperación de documentos digitales, implementando para ello un sistema compuesto por tres módulos: un módulo de conocimiento, un módulo de búsqueda semántica y finalmente un módulo de visualización y navegación de resultados.

PALABRAS CLAVES: Web semántica / Recuperación de información / Documentos digitales / Sistemas multiagentes / Gestión de información / Ontologías

GUZMÁN LUNA, Jaime A.; TORRES PARDO, Durley ; OVALLE, Demetrio A.. SABIOS: una aplicación de la Web semántica para la gestión de documentos digitales. En: Revista Interamericana de Bibliotecología, Vol. 30, No. 1 (ene.-jun. 2007); p. 51-72.

Artículo recibido: 12 de marzo de 2007. Aprobado: 18 de abril de 2007.

Abstract

The Semantic Web seeks to convert information into knowledge. To do so, it must be based fundamentally on semantic and descriptive markings not only of the documents but the data as well, by means of automatically structured and legible information that provides the machines with a greater capacity for managing and retrieving the data. SABIOS proposes to introduce emergent

* Artículo derivado de la investigación: *Desarrollo de un Sistema de Recuperación de Información Semántico Especializado en Artes Plásticas y Hábitat*, Proyecto financiado por la DIME de la Universidad Nacional de Colombia, Sede Medellín.

** Magíster en Ingeniería de Sistemas. Profesor Escuela de Sistemas, Universidad Nacional de Colombia, Sede Medellín, Colombia. jaguzman@unal.edu.co

*** Ingeniera de Sistemas. Auxiliar de Investigación Escuela de Sistemas, Universidad Nacional de Colombia, Sede Medellín, Colombia. durleytp@gmail.com

**** PhD en Ciencias de la Computación. Profesor Escuela de Sistemas, Universidad Nacional de Colombia, Sede Medellín, Colombia. dovalle@unalmed.edu.co

technologies based on semantics and combined with multi-agent systems and information retrieval techniques in order to improve the processes of insertion, cataloging, and retrieval of digital documents, thereby implementing a system composed of three modules: a knowledge module, a semantic search module, and finally a module for visualizing and navigating the results.

KEY WORDS: The semantic Web / Information retrieval / Digital documents / Multi-agent systems / Information management / Ontologies

GUZMÁN LUNA, Jaime A.; TORRES PARDO, Durley; OVALLE Demetrio A. SABIOS: an application of the semantic Web for the management of digital documents. In: Revista Interamericana de Bibliotecología, Vol. 30, No. 1 (jan.-jun. 2007); p. 51-72.

1. Introducción

El creciente número de documentos digitales que surgen cada día al interior de las instituciones universitarias como resultado de la actividad académica e investigativa, así como la necesidad consiguiente de obtener aquellos documentos que responden a la necesidad informativa de los miembros de estas instituciones, ocasiona que los centros documentales que gestionan esta información, se interesen por tener mejores sistemas informáticos que les permitan cumplir con su misión. Es este el caso del centro de documentación “Estación Virtual” adscrito a la Escuela de Artes Plásticas, de la Universidad Nacional de Colombia, sede Medellín. En esta dirección surge el proyecto SABIOS, un sistema Multi-Agente de recuperación semántica [15] de documentos digitales especializados en artes plásticas, el cual fue desarrollado utilizando diferentes técnicas de la Web Semántica[39], disciplina que propone mecanismos de representación de la información mediante el uso de las ontologías [14, 6, 11] para expresar su significado semántico con el fin de que las máquinas logren su entendimiento y así poder llevar a cabo de manera automática la tarea de recuperación de información.

La propuesta presentada en este artículo tiene como objetivo mostrar las bondades de las tecnologías de la Web Semántica en la solución de problemas que se presentan en un sistema informático orientado a la gestión de documentos [38] en un entorno universitario.

Particularmente en SABIOS se plantea una solución a los siguientes problemas: a) En la Escuela de Artes Plásticas no existía una terminología y un modelo de clasificación consensuado entre los diferentes actores que intervienen en los procesos de catalogación y uso de los documentos digitales que ellos mismos generaban, de manera que se posibilitara un lenguaje unificado y facilitar así las labores de cada usuario en la publicación y recuperación de información dentro del sistema, b) En un sistema informático orientado a la gestión de documentos, no es

posible que el usuario pueda conocer las relaciones semánticas entre los diferentes términos de clasificación existentes al interior del sistema, debido a lo cual se le dificulta la formulación de las búsquedas en el sistema [34], c) Cuando se utiliza un sistema de gestión documental, no existe una participación activa de los autores de los documentos en las actividades de almacenamiento y clasificación de los mismos.

A continuación, se describen las principales características del sistema SABIOS y como enfrenta la solución a los anteriores problemas mediante el uso de las tecnologías de la Web Semántica, para lo cual este documento está organizado de la siguiente forma: en la sección 2, se presenta un marco conceptual en relación con los temas de recuperación de información y Web Semántica que soportan el sistema SABIOS; en la sección 3, se presenta una visión general del sistema, detallando que módulos lo componen y que tipos de actores que intervienen en él; en la sección 4, se detallan los aspectos más importantes de la base del contexto, mecanismo mediante el cual se representa y gestiona el conocimiento en el dominio de artes plásticas que maneja el sistema; en la sección 5, se describe la base documental, en la cual se representa y gestiona el conocimiento que caracteriza los documentos digitales; en la sección 6, se describe el proceso de búsqueda y recuperación de documentos mediante el uso de las bases de conocimiento anteriormente descritas; en la sección 7, se explican las principales funcionalidades del módulo de visualización y navegación del sistema, con el cual se pretende mejorar los resultados presentados al usuario final; en la sección 8, se detalla cómo se mantiene la base documental dentro del sistema, en la sección 9, se compara el sistema SABIOS con algunos trabajos relacionados en la literatura y finalmente en la sección 10, se presentan las conclusiones y trabajos futuros.

2. Marco conceptual

Dado que SABIOS utiliza conceptos relacionados con los sistemas de recuperación de información y la Web semántica, se hace necesario revisar estos temas.

La recuperación de información (RI) tiene que ver con la representación, almacenamiento, organización y acceso a los ítems de información [1]. Es una disciplina que trata sobre la localización de una determinada información en una base de datos o repositorio de datos [22]. El objetivo básico de la recuperación consiste en retornar la información relevante, comúnmente documentos digitales, en respuesta a una pregunta formulada por un usuario y ordenarla de acuerdo a su relevancia [26].

Para que un sistema informático pueda cumplir la función de recuperar información, necesita traducir las necesidades del usuario en una consulta que pueda

ser procesada. Un sistema con estas características, debe utilizar un modelo o vista lógica que permita representar los documentos y la consulta; este modelo permite obtener una estrategia para evaluar la relevancia de los documentos respecto a una consulta y establecer su orden en la salida (resultado). La recuperación de información puede ser abordada de acuerdo a la estrategia de búsqueda utilizada, generalmente se habla de dos grandes estrategias [16]: la consulta (*querying*) y la navegación (*browsing*).

En la estrategia de búsqueda por consulta, el usuario ingresa una serie de palabras claves que son comparadas con el espacio documental de forma tal que se retornen los documentos más relevantes acorde a la necesidad de información del usuario. En este tipo de búsqueda se requiere que el usuario tenga claro su necesidad de información y sepa como expresarla mediante el uso de palabras representativas. Esta estrategia es la más utilizada actualmente, tanto en los sistemas de gestión de documentos como en la Web; sin embargo en muchos casos no resulta conveniente debido a que el usuario no tiene claro lo que busca, o no es capaz de formular su necesidad de información a través de palabras, razón por la cual se requiere de alternativas que le faciliten el acceso a la misma.

La estrategia de búsqueda por navegación ofrece esta alternativa, pues el usuario explora el conjunto de documentos sin necesidad de especificar su necesidad a través de palabras. Esta estrategia consiste en explorar visual y espacialmente los documentos con el objetivo de encontrar información de interés.

A pesar de las estrategias anteriormente expuestas, la recuperación de información sigue siendo un problema complejo. Con la aparición de la Web Semántica [3], el concepto de recuperación se ha extendido, hoy es posible realizar el proceso de búsqueda y recuperación de información utilizando metadatos que describen un dominio en particular. Éstos permiten representar los documentos desde una perspectiva semántica y pueden ser utilizados tanto en el proceso de búsqueda por consulta como en el proceso de búsqueda por navegación.

Los metadatos asociados a los documentos son definidos a través del uso de ontologías. Una ontología se puede definir como una especificación explícita y formal sobre una conceptualización compartida [14]. Las ontologías permiten definir de manera compartida y consensuada conceptos sobre un área específica del conocimiento y establecer las relaciones entre dichos conceptos. Esto permite a humanos y máquinas comunicarse de manera eficiente al proveer una definición común de un dominio. Además las relaciones entre conceptos son la base para el razonamiento semántico. Para la construcción de estas ontologías se han desarrollado varios metalenguajes y estándares, entre los que se tiene el RDF [39] y el OWL [27].

Las ontologías en la Web semántica tienen muchas aplicaciones. Una de estas es su utilización en la representación de los Tesauros [12, 25], mecanismos utilizados ampliamente en centros de documentación y bibliotecas para definir los metadatos que permiten la descripción de los documentos. En esta dirección surge el SKOS-Core [35], el cual se ha convertido en una de las principales propuestas para la representación de Tesauros documentales en la Web Semántica.

Una segunda aplicación de las ontologías en la Web Semántica es la visualización y navegación de resultados en un entorno de recuperación de información, la cual da soporte a cuatro tareas básicas: una visión general de la información, un acercamiento, el filtrado de la información y por último una visión detallada de la parte de interés. En este contexto, la visualización de la información ayuda al usuario a comprender y navegar a través del espacio documental y de este modo recuperar los documentos que satisfagan su necesidad de información. Dadas las características de la búsqueda por navegación, la visualización puede ser aplicada a este modelo de búsqueda más fácilmente que en el modelo de búsqueda por consulta. La Web semántica y sus características han permitido adaptar la visualización de la información para que tome ventaja de las propiedades que ofrecen las ontologías [7, 11]. Es aquí donde se habla de *visualización de información basada en ontologías* [10].

3. Metodología propuesta y descripción general de resultados

El desarrollo del proyecto SABIOS se basó en la metodología MAS-CommonKADS [17] la cual da soporte a las etapas de conceptualización, análisis y diseño de un sistema basado en agentes [15]:

- *Fase de conceptualización*: El objetivo de esta fase es la definición y el alcance funcional del proyecto. Los principales resultados de esta fase son la identificación del problema a resolver desde el punto de vista del usuario, la identificación de los actores que interactúan con el sistema y la determinación de los casos de uso que ayudan a entender los requisitos del sistema.
- *Fase de análisis*: Esta fase propone seis modelos para la definición del sistema: agente, tareas, experiencia, coordinación, comunicación y organización.
- *Fase de diseño*: una vez se han definido y descrito todos los elementos que hacen parte del sistema, los agentes, las tareas, la comunicación entre agentes, el entorno y el conocimiento, se define como transformar dichas descripciones en un sistema real, para este caso MAS-Commond KADS propone el modelo de diseño, en donde además de todas las tareas de diseño clásico contempla tareas como el diseño de los agentes, las bases de conocimiento, las caracterís-

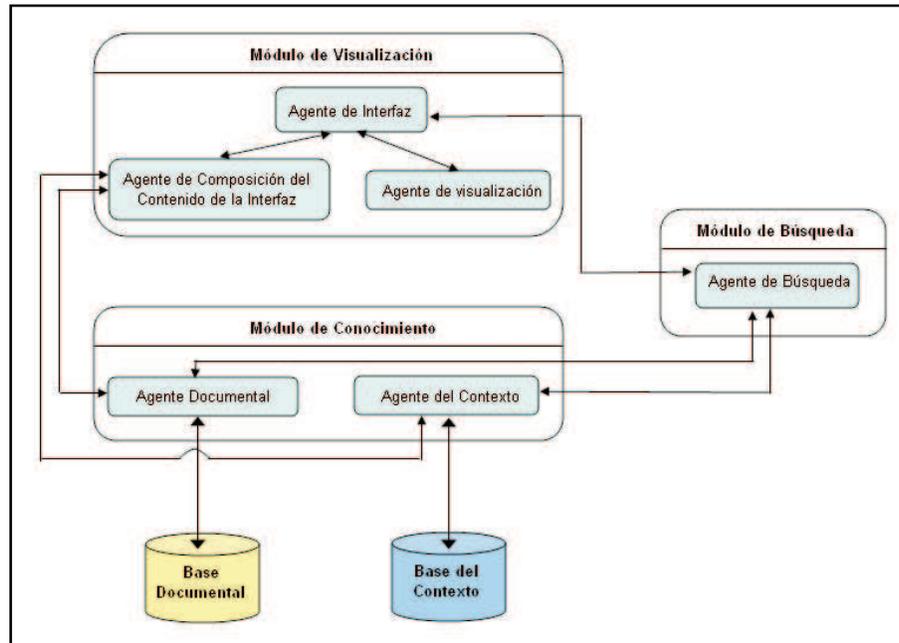
ticas de red (en caso de que los agentes se muevan sobre alguna), y el diseño de los protocolos de comunicación.

Por último, dado que la metodología MAS-Common KADS no describe como se debe realizar la *Fase de Implementación*, para este proyecto en particular se definieron los siguientes aspectos: la implementación de las ontologías del dominio, la construcción de las bases físicas del conocimiento, la creación de los componentes y respectivos agentes y la implementación de la Interfaz para Internet.

Los resultados obtenidos en el proyecto SABIOS se concretan en tres módulos que conforman el sistema que lleva el mismo nombre del proyecto, los cuales tienen como objetivo abordar los problemas mencionados anteriormente (ver figura 1):

- Con el fin de atacar el problema relacionado con la heterogeneidad de los modelos mentales de los actores que intervienen en los procesos de catalogación y uso de los documentos digitales, se implementó el *módulo de conocimiento*, el cual contiene el conocimiento consensuado relacionado con: a) la terminología y el modelo de clasificación del dominio de artes plásticas manejado al interior de la Escuela de Artes de la Universidad Nacional de Colombia y b) la representación de los documentos digitales que maneja el sistema. Este módulo está conformado por el agente documental y el agente del contexto, los cuales se relacionan con sus respectivas bases de conocimiento (base documental y base del contexto).
- Con el fin de solucionar el problema relacionado con el desconocimiento por parte del usuario, de las relaciones semánticas entre los diferentes términos de clasificación al interior del sistema, se desarrollaron los siguientes módulos: a) el *módulo de búsqueda semántica*, el cual realiza el proceso de razonamiento sobre las bases de conocimiento, para realizar la recuperación de los documentos digitales, b) el *módulo de visualización*, el cual le permite al usuario inicialmente expresar su intención de búsqueda mediante una consulta de términos semánticos asociados a los documentos y posteriormente mejorar su búsqueda, mediante la navegación sobre el modelo de clasificación del dominio en artes plásticas definido en el módulo de conocimiento
- Para dar solución al problema de catalogar los documentos bajo una visión más cercana al autor, se le permite a los autores, además de los documentalistas, el mantenimiento de la base documental, mediante tareas específicas del módulo de visualización y del módulo de conocimiento.

FIGURA 1
Modelamiento del sistema SABIOS



En SABIOS, se definieron los siguientes actores potenciales, asociándole a cada uno, de acuerdo con su perfil, las funciones que se consideraron necesarias habilitar en el sistema.

- *Actor usuario*, el cual representa el rol del usuario que acude al sistema a realizar una consulta. A él se le asocian, en primer lugar, la *función búsqueda semántica*: para lo cual ingresa al sistema una lista de conceptos, con el objetivo de que el sistema le retorne instancias de libros asociados a estos, en segundo lugar, la *función búsqueda por metadato [9]*: con la cual le ingresa al sistema un valor de metadato de un documento, tal como título, descripción, autor, etc., para que el sistema realice una búsqueda de documentos que contengan los metadatos especificados.
- *Actor autor*, quien representa a aquella persona o personas que producen un documento. A este actor se le asocian dos funciones: *función ingresar documento*, con la cual el sistema recibe los metadatos que se enlazan al documento y se le asocian los descriptores semánticos (conceptos) registrados en la base de contexto; *función editar documento*: con la cual el usuario autor puede modificar metadatos, pero sólo de aquellos documentos elaborados por él.
- *Actor documentalista*, representa el rol del catalogador de los documentos y administrador del centro documental. A este usuario se le asocian tres funcio-

nes: *función editar documento*, para cuando se quiera modificar algún metadato de un documento existente en el sistema, *función borrar documento*, le permite borrar un documento de la base documental, *función agregar documento*, para insertar un documento a la base documental.

- *Actor administrador*: este representa al administrador del sistema, cuyas funciones están asociadas a *acceder a la base de contexto* para editar la ontología de contexto.

Particularmente en la implementación de SABIOS se utilizaron las siguientes tecnologías: en las ontologías se utilizó OWL en asocio con la herramienta Protegé [28], en el caso de las bases del conocimiento se utilizó la herramienta Jena [19] en asocio con el gestor de bases de datos MySQL [23], en la implementación de los componentes y respectivos agentes se utilizó JADE [5], una plataforma Multi-Agente implementada en JAVA [8] y para el desarrollo de las interfaces para la Web se utilizó JSP [18]. En las siguientes secciones de la 4 a la 8, se describen con más detalle las principales características de este sistema.

4. Representación del conocimiento en el contexto de artes

En la base del contexto se almacena el conocimiento en el dominio de artes. Para su implementación se siguieron los pasos sugeridos en la metodología para la creación de ontologías “*Ontology Development 101*” [24], en donde sus principales recomendaciones consisten en:

1. Determinar el dominio y ámbito de la ontología
2. Determinar la intención de uso de la ontología
3. Reutilizar ontologías o vocabularios controlados existentes.
4. Enumerar los términos importantes del dominio.
5. Definir jerarquía de clases
6. Crear las instancias

Aunque aparentemente la cuestión de determinar el dominio estaba clara, el tema de artes plásticas presenta una amplia categorización de elementos y conceptos, razón por la cual establecer un límite de profundización para cada rama fue una tarea muy cuidadosa, ya que cada una de ellas podía generar por si sola, otra ontología completa. La definición de cada categoría temática, se basó en el plan de estudios del programa curricular de Artes Plásticas de la Facultad de Arquitectura de la Universidad Nacional y la validación y consenso de los términos fue llevada a cabo entre varios docentes expertos en el tema, usuarios del centro documental y el encargado de la catalogación de los documentos. Como resultado, las áreas temáticas en la ontología se limitan a la definición de categorías genéricas por tema y subtemas.

La intención de uso de la ontología era la gestión documental, por ello la recomendación en la que se centro la atención fue precisamente en la reutilización de ontologías existentes. Es así como se acordó la instanciación del dominio bajo un esquema de tesauro, por ello la primera constatación hecha, apuntó a la decisión de reutilizar una ontología que ha sido creada precisamente para modelar conceptos bajo esquemas de tesauros [13, 20, 21]; dicha ontología es el SKOS-Core.

En el SKOS-Core, la idea base del uso de un Schema RDF [32], reside en su capacidad para permitir la definición de conceptos (*skos:Concept*) y esquemas de conceptos (*skos:ConceptSchema*). Un concepto se define como una unidad de pensamiento que puede ser definida o descrita. A su vez un esquema de conceptos no es otra cosa que una colección de conceptos. Cada concepto solo puede tener una etiqueta preferente, -lo que los documentalistas denominan descriptor o término preferente- y un número limitado de etiquetas alternativas, denominado término no-descriptor o no-preferente. La codificación de las etiquetas correspondientes a los términos preferentes y no preferentes pertenecientes a un concepto, se realiza mediante las propiedades *skos:prefLabel* y *skos:altLabel* respectivamente (ver Tabla 1). Esta segunda etiqueta representa la relación de sinonimia o equivalencia entre dos conceptos. Adicionalmente, se contempla el *skos:Hidden*, el cual permite generar el mismo concepto *skos:prefLabel*, pero esta vez con errores ortográficos o hasta de digitación. Adicionalmente, el SKOS-Core permite representar las relaciones básicas entre los conceptos, manejadas por la mayoría de los tesauros mediante la familia *skos:SemanticRelation*, bajo la cual se encuentran las relaciones jerárquicas definidas por las propiedades *skos:narrower* y *skos:broader*, para términos específicos y generales respectivamente, y *skos:related* para términos relacionados (ver Tabla 1).

TABLA 1
Descripción de las propiedades definidas en la Ontología SKOS-Core

Propiedad SKOS-Core	Tipo / Actúa	Definición
<i>Skos: prefLabel</i>	Términos	Preferente o Descriptor
<i>Skos: altLabel</i>	Términos	No preferente o No descriptor
<i>Skos: Hidden</i>	Términos	Definición errada del Descriptor
Relaciones semánticas	Tipo / Actúa	Definición
<i>Skos: narrower</i>	Términos	Términos específicos
<i>Skos: broader</i>	Términos	Términos específicos
<i>Skos: related</i>	Términos	Términos específicos

Como primer paso, en el proceso de diseño se procedió a extraer los términos más relevantes del plan de estudios del programa curricular de Artes Plásticas. El orden jerárquico depende directamente del término artes plásticas (término más general en el esquema de conceptos). En este proceso se definieron ocho categorías principales que cubren las temáticas generales del dominio, donde cada una de ellas representa un concepto. A cada una de las ocho categorías, se le realizó un mapeo de las relaciones de herencia que tenían con otros conceptos. Es decir, se efectuó el recorrido a todo el plan de estudios y se fueron extrayendo otros conceptos que pudieran estar referenciados como conceptos subordinados, con el fin de categorizar de manera precisa cada concepto; para esto se utilizaron las propiedades *skos:broader* y *skos:narrower* definidas en el SKOS:Core.

El proceso anterior, se repitió con cada una de las ocho categorías, generándose un árbol jerárquico por cada una de ellas. Cada categoría del árbol fue validada con un grupo limitado de docentes o estudiantes, que se identificaban por tener trayectoria académica en alguna línea de investigación o por reportar una elevada intensidad horaria de cátedra impartida, además se contó con la opinión del documentalista, esto con el fin de generar consensos y corroborar la correcta selección, jerarquización y catalogación de cada concepto en el árbol constituido.

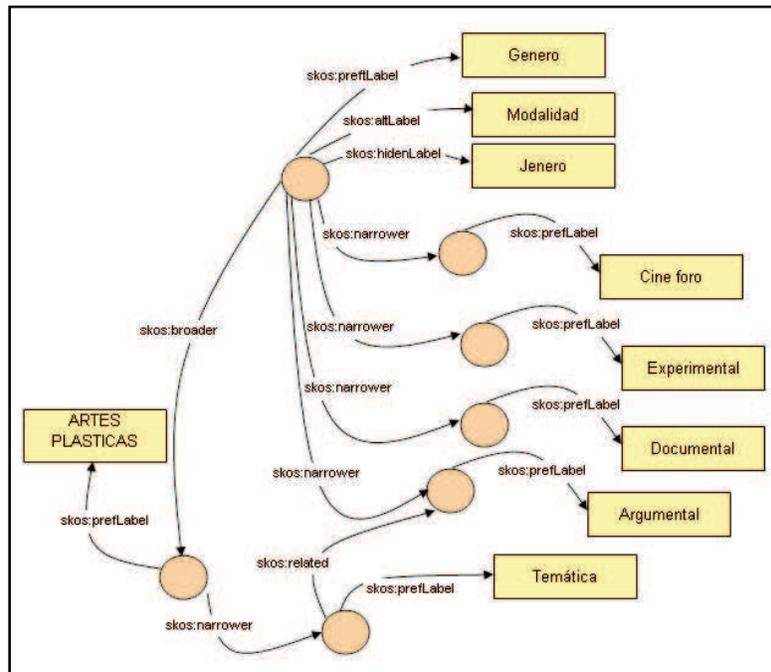
El segundo paso consistió en adicionar relaciones de asociación entre los conceptos, es decir representar como se relacionaba cada concepto con los demás bajo un tipo de relación de asociación, que representa una similitud temática entre los conceptos, diferente a la de jerarquía definida en el paso anterior. Para su implementación se utilizó las propiedades *skos:related* del SKOS-Core.

Como resultado de este proceso, se identificaron e incluyeron 209 conceptos, los cuales conformaron el árbol jerárquico, casi siempre hasta un tercer nivel, sin omitir que algunas ramas contienen más niveles de profundidad. El primer nivel, donde se encuentran los conceptos principales bajo el término Artes, está conformado por ocho descriptores: artes plásticas, temática, soporte, herramienta, material, género, técnica y época. En la elección de los conceptos se tuvo siempre presente la posibilidad de que estos, en algún momento, pudieran ser descriptores de un documento digital. No obstante, siempre queda abierta la posibilidad de extender la ontología a otros conceptos y descripciones, gracias a la propiedad que tienen las tecnologías de la Web semántica de facilitar esta tarea.

Por último, tanto la ontología del SKOS-Core como las instancias que conforman el dominio de artes plásticas, fueron implementadas utilizando el lenguaje OWL [27], y convertidas a RDF-Schema [32] y RDF [4, 32] respectivamente, lenguajes menos expresivos que el OWL, pero que permiten su manipulación con una API llamada JENA [19], un ambiente de trabajo para construir aplicaciones en la Web Semántica. Mediante esta API, la ontología del SKOS:Core y sus instan-

cias se guardaron en una base de datos como MySQL para su posterior manipulación. La implementación de la base de contexto, fue realizada sobre la herramienta Protégé [28]; la figura 2, detalla apartes de las relaciones jerárquicas establecidas para el concepto género, utilizando la ontología de representación SKOS-Core.

FIGURA 2
Representación del concepto género en la ontología SKOS-Core

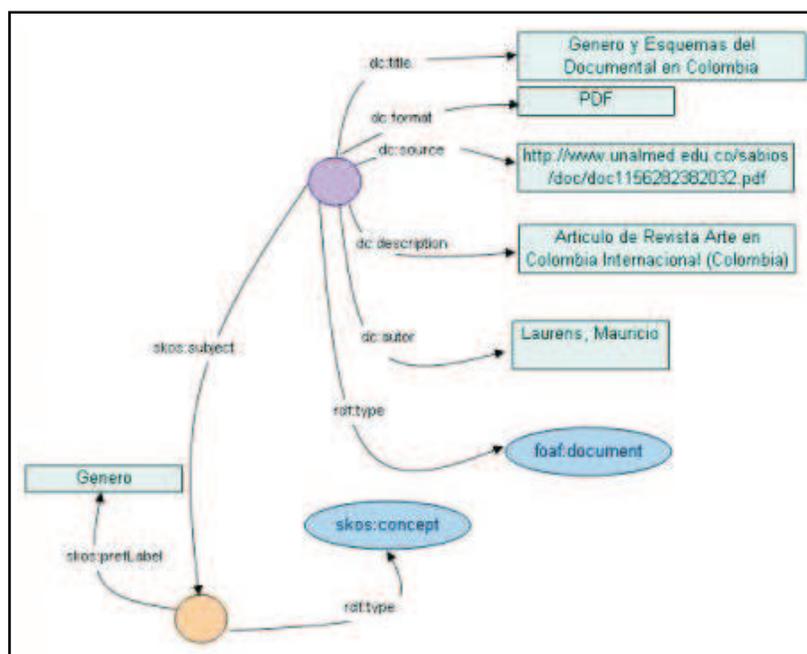


5. Representación de documentos

En la base documental se almacena el conocimiento relacionado con las características de cada uno de los documentos manejados por el sistema. Para esto se implementó una ontología que permite representar los documentos y sus principales propiedades. En la figura 3, se detalla un ejemplo de la instancia “genero” de esta ontología. En la ontología se define la clase *foaf:document* para representar un recurso documento y se le relacionan sus propiedades haciendo uso de algunos términos definidos en la especificación de *Metadatos Dublin-Core* [9], entre los que se encuentran: el título del documento (*dc:title*), el formato (*dc:format*), la dirección donde se encuentra (*dc:source*), una descripción del documento (*dc:description*) y el autor (*dc:autor*). Adicionalmente, a cada uno se le asocia la propiedad *skos:subject*, definida en el SKOS:Core con la que se pretende relacionar a cada documento los conceptos de la base del contexto que lo describen

semánticamente. De esta manera es posible enlazar la Base de Contexto, con la Base Documental, permitiendo que el módulo de recuperación semántico, retorne el título, el autor, la descripción, la fecha de publicación o el idioma de un documento en particular.

FIGURA 3
Representación de un documento en la Base Documental



6. La recuperación semántica de documentos

En el sistema SABIOS, el usuario formula la consulta en forma de términos semánticos (conceptos), mediante el agente interfaz, quien solicita al agente de búsqueda la tarea de encontrar los documentos relacionados con estos términos semánticos. El agente de búsqueda toma los términos semánticos que componen la consulta y solicita al agente de la base del contexto que determine si estos términos coinciden con los conceptos existentes en la base del contexto, para lo cual este agente hace una búsqueda tanto en las etiquetas preferentes (*skos.prefLabel*) como en las etiquetas alternas (*skos.altLabel*) e incluso en las de los errores (*skos.Hidden*) de cada concepto (*skos:concept*). Si al término de la consulta el Agente encuentra que existe uno ó más de estos conceptos, procede a buscar para cada uno de estos, las relaciones semánticas que tenga con otros conceptos (relaciones de herencia - *skos:broader*, y relaciones de generalización

- *skos:narrower* y relaciones de asociación - *skos:related*), con el fin de encontrar otros conceptos que sean relevantes de manera semántica con la consulta hecha por el usuario.

Luego de encontrar estos nuevos conceptos, el agente de la base del contexto retorna al de búsqueda un documento RDF que contiene una lista de tripletas RDF con la siguiente información de cada uno de los conceptos originales y sus asociados: el identificador único del concepto, la propiedad que identifica el tipo de relación semántica que establece con otros (*skos:broader* ó *skos:narrower* ó *skos:related*) y el identificador único del concepto asociado. En caso de que no exista coincidencia entre los términos que conforman la consulta original del usuario y los conceptos existentes en la base del contexto, el agente retorna un fallo en la solicitud.

Cuando el agente de búsqueda recibe del agente de la base del contexto el documento RDF con la lista de conceptos y sus relaciones, solicita al agente de la base documental que busque todos los documentos que estén marcados semánticamente con el conjunto de conceptos que conforman esta lista, para lo cual le envía junto a esta solicitud el documento RDF que contiene tales conceptos. En caso de existir documentos que cumplan con lo solicitado, el agente de la base documental le devuelve al agente de búsqueda un documento RDF con una lista de tripletas RDF que contiene la siguiente información de cada documento encontrado: el identificador único del recurso documento existente (*dc:source*), la propiedad *skos:subject*, que representa el concepto asociado al documento y por último el identificador único del concepto asociado. En caso que el agente documental no encuentre ningún documento, este retorna al agente de búsqueda un fallo.

Cuando el agente de búsqueda recibe del agente documental el documento RDF, con la información de los documentos, este los regresa al agente de interfaz. En caso de recibir respuesta de un fallo, ya sea por parte del agente de la base del contexto o del agente documental, el agente de búsqueda le retorna al agente de interfaz un fallo.

Por último, cabe destacar que el modelo planteado en esta sección utiliza en la implementación del Agente del Contexto y del Agente Documental, un mecanismo de consulta llamado SPARQL [36] que hace parte la herramienta JENA, el cual permite durante las tareas de razonamiento de estos agentes acceder al conocimiento (ontologías y sus instancias) almacenado en las bases de datos que conforman la Base del Contexto y la Base Documental.

7. Visualización y navegación de resultados

El agente de interfaz encargado de interactuar con los usuarios del sistema, luego de recibir los dos documentos RDF (la lista de conceptos y sus relaciones, al

igual que la lista de los documentos asociados a tales conceptos), debe interactuar con los agentes de visualización y de composición de servicios de visualización, con el fin de presentarle al usuario los resultados de una manera apropiada y tener la opción de inferir más información sobre dichos documentos.

En un primer paso, el agente de interfaz le solicita al agente de visualización que defina que información asociada a los documentos que hacen parte de la lista resultante de la consulta, se mostrarán al usuario. Para esto, el agente de interfaz le envía al agente de visualización el documento RDF con la lista.

Para llevar a cabo su tarea, el agente de visualización maneja una ontología de visualización la cual consta de cuatro elementos básicos: el dominio que describe cuales elementos serán visualizados, la representación que define cómo serán visualizados, la fuente que describe de dónde se sacan las instancias de los elementos, y por último, la visualización que permite unificar el dominio, la representación, y la fuente, en un solo servicio de visualización. Es así como, de manera dinámica, este agente genera para cada documento existente en el documento RDF, una tripleta que hará parte de una instancia temporal de esta ontología en la cual se indica como se extraerá la información de la base documental. En la Tabla 2, se muestran apartes de una instancia de la ontología de visualización para un documento específico. La implementación de esta ontología se realizó mediante el Protegé, utilizando el RDF-Schema, la cual se optó para guardar en un archivo de texto en vez de una base de datos.

Luego de generada la nueva instancia de esta ontología, se le retorna al agente de interfaz en un archivo RDF que llamamos *archivo de visualización*. Con esta información, el agente de interfaz procede a solicitarle al de composición de interfaces que construya la interfaz en la que se mostrará al usuario la información contenida en el archivo de visualización. Para esto, el agente compositor genera una interfaz en HTML, donde además de la información que se desea mostrar, se incluyen los elementos necesarios para que el usuario interactúe con la interfaz.

En la figura 4, se presenta la interfaz, que señala los resultados a una consulta realizada textualmente, por el concepto género. Los resultados devueltos por el buscador semántico, sugieren que existen dos documentos dentro de la base documental, los cuales hacen referencia a ese concepto; sin embargo, esa relación no es directa, es decir, son los conceptos más específicos al concepto género, los que se relacionan directamente con la búsqueda textual, en otras palabras, la búsqueda retorna todos los documentos que guardan relación con los conceptos más específicos a género, adicionalmente la interfaz, señala al lado de cada concepto, y entre paréntesis, la cantidad de documentos con los que se relaciona cada concepto específico.

Adicionalmente, la interfaz permite la opción de realizar expansión de búsqueda la cual consiste en navegar a través de la red semántica de conceptos correspondientes al dominio, esta facultad permite seleccionar conceptos y aportar a la necesidad de información del usuario. Esta relación puede ser del tipo específica, referente, o general:

- Términos específicos: indica que el concepto es mas específico que el de búsqueda, es decir, el concepto de búsqueda hace referencia a una idea más general, mientras que el concepto específico solo trata o incluye una parte del concepto de búsqueda. Ej. documental (concepto específico), género (concepto de búsqueda).
- Términos referentes: especifica una relación semántica entre conceptos del tipo asociativo. Indica que un concepto referente tiene algún vínculo con el concepto de búsqueda. Ej. cine (concepto referente), género (concepto de búsqueda).
- Términos generales: muestra una relación semántica entre conceptos del tipo general. Indica que el concepto es más general que el de búsqueda, es decir, el concepto de búsqueda esta incluido dentro del concepto mas general (refiriéndose al significado). Ej. artes (concepto general), genero (concepto de búsqueda).

FIGURA 4
Interfaz de resultados correspondiente a la visualización y navegación de la búsqueda textual por el concepto “género”



8. Mantenimiento de la base documental

La base documental también debe ser preservada y actualizada constantemente, para realizar esta tarea el sistema habilita tanto al autor como al documentalista, para registrar, editar o eliminar los documentos en la base de datos cuando se considere necesario. Para realizar esta tarea, SABIOS, cuenta con registros y password que habilitan las funcionalidades de cada uno de los actores que intervienen la base documental, esto con el fin de asegurar derechos de autor y la integridad de los datos entre otros.

En la figura 5, se muestra la interfaz presentada al documentalista y al autor, en la que puede navegar en la base de contexto, con el fin de facilitarle la tarea de ubicar bajo el concepto más relevante su documento, catalogarlo bajo un concepto y atributos previamente modelizados y consensuados entre los diferentes actores, mitigando así los problemas de divergencia en los modelos conceptuales relacionados con el dominio específico de artes.

FIGURA 5
Interfaz de usuario agregar documento

The screenshot shows the 'Subir Nuevo Documento' (Upload New Document) interface. On the left is a navigation menu with options like 'Opciones Administrador', 'Administrar Tesauro', 'Administrar Base Documental', 'Agregar Documento', 'Editar Documento', 'Administrar Autores', and 'Log out'. The main area contains a form with the following fields: 'Titulo:' (text input), 'Descripcion:' (text area), 'Autores: [agregar]' (text input), 'Fecha publicacion:' (calendar icon), 'Idioma:' (dropdown menu set to 'Español'), 'Formato:' (dropdown menu set to 'PDF'), and 'Temas:' (a list of five empty text boxes, each with an 'editar' button to its right). Below these is a 'Ubicacion del documento:' field with an 'Examinar...' button. At the bottom of the form is a 'Subir Documento' button. The page header includes the Universidad Nacional de Colombia logo and 'SABIOS Sistema de Búsqueda Semántica'. The footer contains contact information and logos for 'un digital', 'universio', and 'GOBIERNO DE NARIÑO'.

9. Trabajos relacionados

Existen varios proyectos donde se incluye la creación de una ontología de visualización, entre estos se tiene el proyecto Neptuno [29] y el proyecto Portal Semántico para Asuntos Internacionales [2].

El proyecto Neptuno tiene como objetivo aplicar técnicas de la Web Semántica y tecnologías basadas en el diseño centrado en el usuario [31] para mejorar los procesos de creación, mantenimiento y exploración de la hemeroteca de un diario digital [30]. Dentro de la plataforma que se plantea en Neptuno, se define un módulo de visualización y navegación de contenidos basados en ontologías que permite definir vistas [33] sobre la ontología de dominio (información periodística). Estas vistas se definen de acuerdo a criterios de usabilidad y estéticos permitiendo presentar de manera legible el modelo semántico subyacente. Este proyecto se diferencia de SABIOS, en que el dominio de aplicación es diferente, mientras en Neptuno es una hemeroteca de un diario digital, en SABIOS se busca la gestión de documentos digitales en un centro documental universitario especializado en artes plásticas. En cuanto al tema de navegación y visualización de la información ambos desarrollan una ontología de visualización y navegación con el fin de mejorar la intención de búsqueda del usuario.

El proyecto Portal Semántico para Asuntos Internacionales se basa en el mismo concepto utilizado por el Proyecto Neptuno y SABIOS. La ontología de visualización contiene los mismos conceptos, pero las vistas son definidas sobre una ontología de dominio diferente (información sobre asuntos internacionales).

A pesar de sus similitudes, el proyecto SABIOS se distingue de los anteriores en varios aspectos. El primero, como se ha mencionado, es el dominio de aplicación el cual gira alrededor de la recuperación de documentos asociados a un proyecto curricular específico en artes plásticas. Un segundo aspecto es que en SABIOS se busca la participación activa de los autores de los documentos para su almacenamiento y catalogación, lo cual no es tenido en cuenta en los anteriores trabajos.

10. Conclusiones y trabajos futuros

En este artículo, se desarrollo un sistema Multi-Agente el cual permite realizar búsquedas semánticas sobre un dominio específico; el sistema desarrollado realiza inferencia sobre la base de conocimiento para encontrar relaciones entre los conceptos definidas en la ontología del sistema, además, introduce la utilización de ontologías para la visualización y representación de un dominio y se señalo como su taxonomía representa un mecanismo para explotar la información semántica

subyacente. La implementación de tal módulo de visualización aporta un valor agregado a SABIOS al extender la búsqueda por consulta de conceptos semánticos a un tipo de consultas por navegación mediante el uso de técnicas de visualización de información que ayudan al usuario a encontrar nuevos conceptos asociados a los conceptos de su intención de búsqueda.

El trabajo desarrollado hasta aquí, representa una experiencia de aplicación de las tecnologías de la Web Semántica en un escenario real, y aporta contribuciones en aspectos como: funcionalidades de búsqueda semántica y exploración de información; desarrollo de una interfaz de usuario capaz de interactuar con una base de conocimiento en un dominio específico y se convierte en una propuesta concreta, para trasladar sistemas de gestión tradicional, como lo son ciertos centros documentales, a plataformas basadas en semántica.

Además de estas ventajas inmediatas el trabajo permite abordar, en un futuro inmediato, sobre las bases implementadas hasta el momento, otros problemas de mayor nivel, por ejemplo, la inclusión de nuevas fuentes (otros centros documentales), géneros de contenidos (imágenes, audio, video), la extensión de la base de representación del conocimiento a otros dominios e idiomas (inglés). Además de mejorar, lo desarrollado hasta aquí, a través de algoritmos de *ranking* que permitan aumentar la efectividad en los resultados presentados.

Aunque SABIOS fue construido bajo un dominio muy limitado como lo es el programa curricular de la Escuela de Artes de la Facultad de Arquitectura de la Universidad Nacional de Colombia, sede Medellín, este puede ser extendido gracias a la plataforma sobre la que ha sido desarrollado, lo que se convierte en una ventaja indirecta del trabajo realizado.

Referencias bibliográficas

1. BAEZA-YATES, R. y RIBEIRO-NETO, B. *Modern Information Retrieval*. New York: Addison-Wesley, 1999, p. 93 – 112.
2. BLÁZQUEZ, M.; LOSADA, S.; SALLA, R.; *et al.* A semantic portal for the international affairs sector. En: *14th International Conference on Knowledge Engineering and Knowledge Management (EKAW'04)*, 5-8 October 2004.
3. BERNERS-LEE, T; HENDLER, J.y LASSILA, O. The semantic Web. *Scientific America*, May. 2001, vol. 284, no. 5, p. 34-43.
4. BRICKLEY, D. y GUHA, R.V. *Resource Description Framework (RDF) Schema Specification 1.0: 3C Working Draft* [en línea] 2002. Disponible en Internet: <http://www.w3.org/TR/rdf-schema> [Consulta: 13 de noviembre de 2006].

5. BELLIFEMINE, Fabio, et al. JADE - A Java Agent Development Framework. Multi-Agent Programming En: Rafael H. Bordini, et al, editores. *Multi-Agent Programming: Languages, Platforms and Applications. Multiagent Systems, Artificial Societies, and Simulated Organizations. s.l. DBLP Books, 2005*
6. CONTRERAS, J. et al. Duontology: an Approach to Semantic Portals based on a Domain and Visualization Ontology. En: 14th International Conference, EKAW 2004, Whittlebury Hall, UK, October 5-8, 2004.
7. CHEN, C. Information Visualization Versus the Semantic Web. En: Visualizing the Semantic Web. [en línea] Londres. Springer, 2002, p. 15-35 Disponible en Internet: <http://www.pages.drexel.edu/~cc345/papers/papers.html> [Consulta: 20 de junio de 2007].
8. DEITEL, M. Harvey y DEITEL J. Paul. *Como programar en JAVA*. Estados Unidos: Pearson Prentice Hall, 2004.
9. *Dublín-Core*. [en línea] Disponible en Internet: <http://es.dublincore.org/> [Consulta: 13 de febrero de 2007]
10. FLUIT, C.; SABOU, M.; HARMELEN, F van. Supporting User Tasks through Visualization of Light-weight Ontologies. En: S. Staab y R. Studer, editores. *Handbook on Ontologies*. Berlin: Springer-Verlag: 2004, p. 415–434
11. GEORGIEVA, R. Ontology-Based Information Representation. *Joint Advanced Student School (JASS). Course 6: Next-Generation User-Centered Information Management*. St. Petersburg - Wednesday, Mar. 30 - Apr. 9 de 2005. [en línea] Disponible en Internet: <http://www14.in.tum.de/konferenzen/Jass05/courses/6/index.html>. [Consulta: 20 de abril de 2007]. (2005).
12. GIL URDICIANI, Blanca. *Manual de lenguajes Documentales*. Madrid: Noesis DL, 2006, p. 269
13. GÓMEZ-PÉREZ A., M. Fernández-Lopez, A. de Vicente Towards a method to conceptualize domain ontologies. En: *ECAI'96 Workshop on ontological engineering*, Budapest, Hungary: s.i. 1996. p. 41-52
14. GRUBER, T.R. Toward principles for the design of ontologies used for knowledge sharing. En: *Formal Ontology in Conceptual Analysis and Knowledge Representation*. The Netherlands: Kluwer Academic Publishers, 1993.
15. HENDLER, James. Agents and the Semantic Web. *IEEE Intelligent Systems*, Mar. – Abr. 2001, p. 30-37.

16. HERRERO-SOLANA, V.y HASSAN, Y. Metodologías para el desarrollo de Interfaces Visuales de recuperación de información: análisis y comparación. *Information Research*, 2005, vol. 11, no. 3, p. 258.
17. IGLESIAS, C. A methodological proposal for multiagent systems development extending CommonKADS. En: ECAI'96 Workshop on ontological engineering; Budapest, Hungary, 1996.
18. *JAVA* [en línea] Disponible en Internet: <http://java.sun.com/products/jsp/> [Consulta: 13 de febrero de 2007]
19. *JENA* [en línea] Disponible en Internet: <http://jena.sourceforge.net/> [Consulta: 13 de febrero de 2007]
20. LÓPEZ-HUERTAS, M. J. Potencialidad evolutiva del tesoro: hacia una base de conocimiento experto. En: *La representación y la organización del conocimiento en sus distintas perspectivas: su influencia en la recuperación de la información*. Actas del IV Congreso ISKO-España EOCONSID99, Granada 22-24 de abril de 1999. Granada: ISKO-Facultad de Biblioteconomía y Documentación, 1999, p. 133-140.
21. MARTÍN GALÁN, B. y RODRÍGUEZ MATEUS, D. Estructuración de la información mediante XML: un nuevo reto para la gestión documental. En: *La Gestión del Conocimiento: retos y soluciones de los profesionales de la información*. VII Jornadas Españolas de Documentación. Bilbao: FESABID, 2000, p.113-123.
22. MEADOW, C. T. *Text information retrieval Systems*. San Diego: Academic Press, 1993.
23. *MySQL*. [en línea] Disponible en Internet: <http://www.mysql.com/> [Consulta: 13 de febrero de 2007]
24. NOY, N.F. y MCGUINNESS, D.L. *Ontology Development 101: A guide to creating your first ontology*. [en línea] Disponible en Internet: <http://www.Ksl.stanford.edu/people/dim/papers/ontology-tutorial-noy-mcguinnessabstract.html> [Consulta: febrero de 2007]
25. MOREIRO GONZÁLEZ, J. A. [et al.]. Mapas conceptuales, topic maps y tesauros. En: *I Jornadas de Tratamiento y Recuperación de Información (JOTRI)* Valencia. 4 y 5 de Julio de 2002. Madrid. Universidad Carlos III.
26. PEREZ-CARBALLO, J. y STRZALKOWSKI, T. Natural Language information retrieval: progress report. *Information Processing and Management*, 2000, no. 36, p. 155-178.

27. *OWL. Web Ontology Language (OWL). W3C Recommendation 10 Feb 2004.* [en línea] Disponible en Internet: <http://www.w3.org/2004/OWL/> [Consulta: 3 de marzo de 2007] *Protégé* [en línea] Disponible en Internet: <http://protege.stanford.edu/plugins/owl/> [Consulta: 7 de febrero de 2007]
28. *PROTÉGÉ.* [en línea] Disponible en Internet: <http://protege.stanford.edu/plugins/owl/> [Consulta: 22 de febrero de 2007]
29. CASTELLS, P. *et al. Neptuno: Semantic web technologies for a digital newspaper archive* [en línea] 1st European Semantic Web Symposium (ESWS 2004). Disponible en Internet: <http://griho.udl.es/publicacions/2004/neptuno-esws04.pdf> [Consulta: 13 de junio de 2007]
30. PULIDO E. *et al.* Newspaper archives on the semantic web. En: NAVARRO-PRIETO, Raquel y LORÉS-VIDAL, Jesús, editores. *HCI related papers of Interacción*. Springer Verlag, 2006.
31. RANDOLPH G.; Bailey Robert; Molich Rolf; Dumas Joe; Spool Jared M. Redish, Janice; Bias. Usability in practice: Formative usability evaluations - evolution and revolution. CHI2002, Minneapolis, 2002.
32. *RDF. RDF Semantics. W3C Recommendation 10 February 2004* [en línea] Disponible en Internet: <http://www.w3.org/TR/rdf-mt/> [Consulta: 13 de febrero de 2007]
33. RICHARD V. *et al.* Duontology: an approach to semantic portals based on a domain and visualization ontology. En: *Martos: International Affairs Portal: A Semantic Web Application. ECAI Workshop on Application of Semantic Web Technologies to Web Communities*, 2004.
34. SALTON, G. y MCGILL, J. *Introduction to modern information retrieval*. New York: McGraw-Hill, 1983, p. 115-28
35. *SKOS-Core* [en línea] Disponible en Internet: http://www2.ub.es/bid/consulta_articulos.php?fichero=13perez2.html [Consulta: 13 de abril de 2006]
36. *SPARQL (2006)* [en línea] Disponible en Internet: <http://www.w3.org/TR/rdf-sparql-query/> [Consulta: 21 de enero de 2007]
37. *XML* [en línea] Disponible en Internet: <http://www.xmlnews.org> [Consulta: 21 de enero de 2007]
38. VALLET, D.; FERNANDEZ, M. y CASTELLS, P. An Ontology Based Information retrieval Model 2nd European Semantic Web Conference (ESWC 2005). *Lectures Notes in Computer Science*, 2005, vol. 3532, p. 455-470.
39. *W3C* [en línea] Disponible en Internet: <http://www.w3.org/TR/RDF/rdf-premier> [Consulta: 13 de enero de 2007]

