

Desarrollo de un Módulo de Visualización y Navegación de Información Basada en Ontologías para el Sistema de Recuperación Semántica SABIOS

Juan David Giraldo, Jaime Alberto Guzmán y Juan Carlos Ceballos

UNIVERSIDAD NACIONAL DE COLOMBIA, Medellín. Escuela de Sistemas.
{jdgiral, jaguzman, jcceballo} @unal.edu.co

Recibido para revisión May-2006, aceptado Jun-2006, versión final recibida Jun-2006

Resumen: Este artículo presenta el desarrollo del Módulo de Visualización y Navegación del Sistema de Recuperación Semántica SABIOS (Sistema de Almacenamiento y Búsqueda de Información Organizada Semánticamente). Este módulo permite dar soporte a las tareas de recuperación añadiendo elementos visuales que faciliten la localización de información relevante para el usuario. Esto implica un cambio en el paradigma de búsqueda, al incorporar la utilización de técnicas de visualización que aprovechen el contenido semántico expresado a través del uso de ontologías. Para el desarrollo del módulo fue necesaria la construcción de una Ontología de Visualización que es manejada por Agentes de Software para definir servicios de visualización. Dentro de las funcionalidades que ofrece el módulo están desplegar los resultados de una consulta, realizar un filtrado de los resultados y expandir una búsqueda con nuevos términos. Esta última se realiza mediante una navegación gráfica sobre la ontología.

I INTRODUCCIÓN

Con el crecimiento de la Web en los últimos años se ha evidenciado un aumento en el volumen de información disponible. Esto acarrea problemas a la hora de buscar información relevante y hace necesario el desarrollo de herramientas que permitan realizar una búsqueda de manera rápida y eficiente sobre este volumen de información. Sin embargo, estas herramientas, también conocidas como buscadores, se encuentran limitadas a la hora de realizar una búsqueda, debido a que utilizan métodos basados en palabras claves [Baeza-Yates y Ribiero-Nieto (1999)] para determinar la información relevante; en esta forma de búsqueda el usuario ingresa una serie de palabras claves que conforman la consulta, ésta es comparada con el espacio documental de modo que se retornen los documentos más relevantes acorde a la necesidad de información del usuario. En ningún momento se utiliza el contenido semántico de los documentos para determinar cuáles son relevantes, y esto se debe a que en la Web actual no existen mecanismos para marcar semánticamente los documentos.

Debido a estas limitaciones, surge la Web Semántica [Berners-Lee, Hendler y Lassila (2001)], la cual puede ser vista como una extensión de la Web actual donde la información recibe un significado mediante el uso de metadatos (datos sobre los datos) [Shah (2003)]. Estos

metadatos pueden ser procesados por agentes de software, permitiendo encontrar la información que es relevante para el usuario utilizando su significado semántico [Shah (2003), Guha, McCool y Miller (2003)].

Uno de los mecanismos para definir estos metadatos son las ontologías, estas permiten especificar un vocabulario controlado (en este caso los metadatos) que sirve para marcar semánticamente un documento. Una ontología se puede definir como *una especificación explícita y formal sobre una conceptualización compartida* [Gruber (1995)]. Las ontologías en la Web Semántica permiten definir de manera compartida y consensuada conceptos sobre un área específica del conocimiento y establecer las relaciones entre dichos conceptos. Para representar estos conceptos y sus relaciones semánticas se han desarrollado varios lenguajes y estándares, entre los que se tiene RDF(S) (*Resource Description Framework*) [Consortium (2004a)] y OWL (*Ontology Web Language*) [Consortium (2004c)]. Estos lenguajes son recomendaciones de la W3C (*World Wide Web Consortium*) [Consortium (2004b)], uno de los principales promotores de la Web Semántica.

Por otro lado, asociado al proceso de búsqueda, está la presentación al usuario de los resultados de manera legible y comprensible. Es aquí donde es útil la visualización de información. Con la aparición de la Web Semántica, y por ende, de la recuperación semántica, surge la

posibilidad de utilizar diferentes técnicas de visualización de información que aprovechen los metadatos [Chen (2002)] y permitan enriquecer el proceso de búsqueda y recuperación. Es aquí donde hablamos de Visualización de información basada en Ontologías [Fluit, Sabou y van Harmelen (2003), Fluit y Wester (2002)].

La Visualización de Información basada en Ontologías puede ser abordada desde dos puntos de vista: la utilización de una ontología de dominio como mecanismo para estructurar la información semánticamente y posteriormente obtener una representación visual de dicha estructura, o la utilización de una ontología auxiliar, conocida como ontología de visualización [Contreras (2003)], que permita describir el proceso de visualización de la información marcada semánticamente a partir de la ontología de dominio.

Por lo tanto, el objetivo de este artículo es presentar el desarrollo de un módulo que utiliza técnicas de visualización de información que aprovechan el contenido semántico de los resultados de una búsqueda expresado a través de ontologías, para enriquecer el proceso de recuperación de información.

Este artículo se encuentra organizado de la siguiente forma. En la Sección 2 se plantea el problema que motiva el desarrollo de un módulo de visualización y navegación. Algunas aplicaciones que utilizan técnicas de visualización de información y ontologías son mostradas en la Sección 3. Luego en la Sección 4 se habla sobre los requerimientos del sistema y la metodología utilizada. Finalmente en la Sección 5 se muestra el prototipo desarrollado con sus características y por último se presentan algunas conclusiones y se plantean trabajos futuros.

2 DEFINICIÓN DEL PROBLEMA

Con el objetivo de aplicar técnicas de la Web Semántica al proceso de búsqueda de información, surge el proyecto SABIOS [Guzmán, Sánchez y D. (2005)], un sistema Multi-Agente de recuperación semántica especializado en los temas de Hábitat y Artes desarrollado en la Universidad Nacional de Colombia, sede Medellín.

Actualmente el sistema SABIOS permite la recuperación de documentos utilizando conceptos que se encuentran definidos en una ontología y que son utilizados para describir los documentos. De este modo la búsqueda se realiza sobre los conceptos a los que se refiere el documento y no por frecuencia de palabras. Estos conceptos se relacionan con otros conceptos a través de una jerarquía, permitiendo definir relaciones de sinonimia, antonimia, asociación, entre otras. Estas relaciones se pueden obtener gracias al uso de ontologías, las cuales representan a través de un vocabulario el conocimiento sobre un dominio específico. Para el caso de SABIOS, se desarrolló una ontología especializada en Hábitat y Artes Plásticas.

Sin embargo, el proceso de recuperación que plantea SABIOS no incorpora elementos que exploten el contenido semántico para el despliegue de los resultados al usuario. Los documentos retornados no están estructurados de acuerdo a la semántica utilizada para su búsqueda, esto significa que, a pesar de utilizar conceptos y relaciones semánticas para la recuperación, esta información no es mostrada al usuario. Además de esta falta de estructura en los resultados, no existe un mecanismo de navegación y exploración de la taxonomía de la ontología, que permita realizar una abstracción por parte del usuario del dominio de conocimiento que está explorando a través de consultas textuales; a partir de una palabra dada por el usuario, la navegación sobre la ontología permite encontrar palabras relacionadas que pueden llevar al usuario a información adicional que no había sido tenida en cuenta inicialmente. Estas falencias hacen que la búsqueda se encuentre limitada ya que no se está aprovechando todo el contenido semántico que provee la ontología.

Es por esto que se pensó en el desarrollo de un módulo de visualización y navegación que utilice el contenido semántico que provee la ontología para inferir más información acerca de los resultados. Este módulo debe permitir la presentación de los resultados de manera amigable, permitiendo definir vistas sobre la estructura semántica subyacente a los resultados. De este modo se separa la capa de contenido y la capa de presentación.

3 ESTADO DEL ARTE

Existen varios proyectos que incluyen la creación de un módulo de visualización que utiliza ontologías para generar representaciones visuales de la información, entre estos se tiene Neptuno [Castells (2004)] y el Portal Semántico para Asuntos Internacionales [Contreras (2004)]. Estos proyectos definen una ontología de visualización que permite separar la estructura de los datos de su presentación.

Por otro lado se han desarrollado diferentes aplicaciones para la recuperación de información utilizando contenido semántico. Entre estas aplicaciones se tiene Kartoo [KARTOO (2004)], Grokker [MIII Groxis (2004)] y Ontogator [Hyvonen, Saarela y Viljanen (2003)]. Estas aplicaciones utilizan técnicas de visualización de información que permiten presentar los resultados al usuario de una forma más comprensible y asequible; una de estas técnicas aplicadas a la recuperación es el Cluster Map [Fluit, Sabou y van Harmelen (n.d.)]. Otro trabajo que incluye elementos visuales para soportar el proceso de recuperación es Matriz Brower [Ziegler, Kunz y Botsch (2002)], que permite la navegación y exploración de los resultados de la búsqueda.

4º MÉTODO PROPUESTO

Dentro del sistema SABIOS es necesario desplegar los resultados de una consulta al usuario y ofrecer la posibilidad de realizar una navegación sobre los conceptos de la ontología para ampliar la consulta. Para realizar esto, se necesita un módulo de visualización y navegación que cumpla con estos requerimientos; este módulo debe componer las interfaces necesarias para capturar la consulta del usuario y desplegar la lista de resultados.

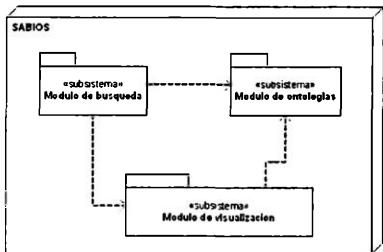


Figura 1: Arquitectura de SABIOS

. Los resultados de la búsqueda son proporcionados por el módulo de búsqueda que hace parte del sistema SABIOS, estos resultados se componen de una lista de documentos en formato PDF (*Portable Document Format*). Además de desplegar los resultados, el módulo de visualización y navegación debe permitir al usuario realizar un filtrado de éstos, este filtrado se realiza con los metadatos utilizados para marcar semánticamente a un documento (e.g., título, autor, tema); éstos son proporcionados por el módulo de ontologías que también hace parte de SABIOS.

Por otro lado, para aprovechar la taxonomía que provee la ontología, el módulo debe permitir al usuario navegar, bien sea textualmente o gráficamente, en esta red semántica de conceptos. Estos conceptos pueden ser utilizados para expandir la consulta del usuario, es decir, agregar estos conceptos encontrados en la ontología a los términos que componen la consulta del usuario. De este modo se le asiste al usuario sugiriéndole nuevos términos. Los conceptos y relaciones de la ontología son provistos por el módulo de ontologías.

Para el desarrollo del sistema se utilizó el paradigma de agentes dado que la arquitectura de SABIOS está pensada como un Sistema Multi-Agente. Se siguió la metodología MAS-CommonKADS [Iglesias (1996)] que da soporte a las etapas de conceptualización, análisis y diseño de un sistema basado en agentes. Para la implementación se utilizó JADE (*Java Agent DEvelopment Framework*), que es una Plataforma Multi-Agente implementada en Java.

5 RESULTADOS

Se desarrolló un prototipo con las características necesarias para dar soporte a la recuperación de información basándose en técnicas de visualización y la utilización de ontologías definidas en lenguaje OWL. El Sistema desarrollado está compuesto de dos módulos, el primero que maneja las solicitudes de visualización y el segundo que se encarga de la composición de interfaces de usuario.

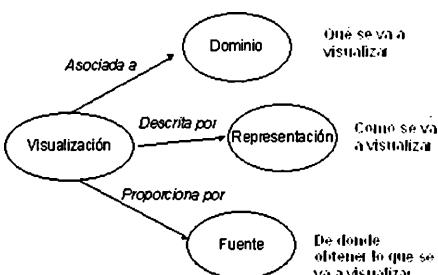


Figura 2: Ontología de visualización

El primer Módulo de Servicios de Visualización contiene un Agente Servicio-Visualización encargado de procesar los servicios de visualización, que son expresados utilizando un vocabulario creado para tal fin. Este vocabulario de definición de servicios de visualización es provisto por una Ontología de Visualización (ver Figura 2). Para describir el servicio, la Ontología provee cuatro elementos: Dominio, Representación, Fuente y Visualización. En el Dominio se describe qué elementos serán visualizados, en la Representación se define el cómo serán visualizados, en la Fuente se describe de dónde sacar 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.

El segundo Módulo de Composición de Interfaces permite dar soporte a la representación de los servicios obtenidos en el primer módulo. Contiene un Agente Compositor que se encarga de llevar la representación de los servicios a una interfaz en HTML, además incluye en la interfaz los elementos necesarios para que el usuario interactúe con el servicio.

Además el sistema incluye un Agente Usuario, encargado de interactuar con los usuarios del sistema, su función es satisfacer la necesidad de información de un usuario, además es el encargado de solicitar el servicio de búsqueda dentro del sistema.

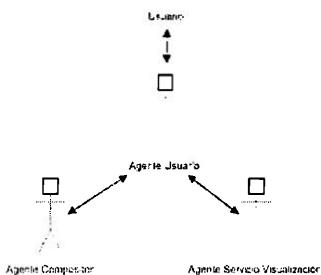


Figura 3: Estructura organizacional

El proceso de visualización de resultados comienza cuando el usuario solicita una búsqueda la cual es atendida por el Agente Usuario, éste se comunica con el módulo de búsqueda enviándole la consulta del usuario en los términos de la ontología de dominio manejada por el sistema (hablado ó artes), luego el Agente Usuario recibe los resultados de la consulta. Estos resultados están representados en un modelo RDF [Consortium (2004a)], el cual contiene los metadatos asociados a cada documento considerado relevante para el usuario.

Los resultados junto con su descripción de visualización (realizada con el vocabulario de la ontología de visualización) son entregados al Agente Servicio Visualización, el cual obtiene la representación en XML de los resultados y las características estéticas. Luego el Agente Compositor construye una página HTML con los resultados y los elementos necesarios para interactuar con ellos. Entre estos elementos está la posibilidad de realizar un filtrado de los resultados o realizar una expansión de la búsqueda.

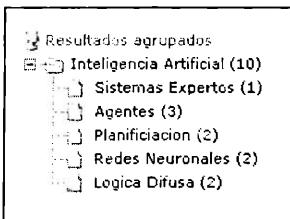


Figura 4: Interfaz resultados agrupados

El filtrado consiste en un agrupamiento de los documentos retornados utilizando los metadatos con los que se marcaron semánticamente los documentos (ver Figura 4). Un ejemplo de este filtrado puede ser por término,

autor, tema, fecha de publicación, etc. La expansión de la búsqueda consiste en agregar nuevos términos a la consulta del usuario, estos términos se obtienen a través de una navegación gráfica del usuario sobre la ontología de dominio (ver Figura 5).

Particularmente, el módulo de visualización y navegación utilizan la técnica de visualización en red y jerárquica.

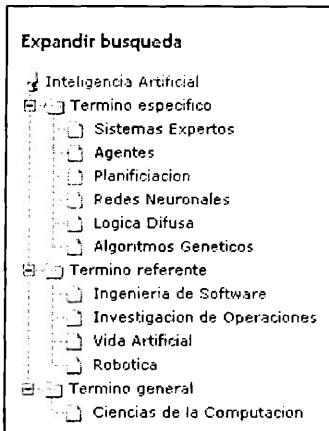


Figura 5: Interfaz Expansión de búsqueda

De este modo se diferencia de las aplicaciones referenciadas en la Sección 3, en el sentido de proveer diferentes vistas de la misma información semántica. Por lo tanto se puede mostrar los resultados agrupados de forma jerárquica, similar a como se hace en Matrix Brower y a su vez se puede mostrar las relaciones semánticas del concepto de búsqueda a través de una visualización en red, similar a como lo realiza Kartoo y Grokker.

6 CONCLUSIONES Y TRABAJOS FUTUROS

El desarrollo de un Módulo de Visualización y Navegación para el sistema SABIOS, ha permitido dar soporte a la tarea de recuperación de información, mejorando la experiencia del usuario al utilizar el contenido semántico de los documentos para generar una representación visual de los resultados, que facilita la localización de los documentos relevantes. Esto implica la aplicación de tecnologías emergentes en la visualización, como la utilización de ontologías para el marcado semántico de información y como mecanismo para estructurar los resultados de una búsqueda.

En este artículo se han presentado los conceptos relacionados con la visualización de información en la Web Semántica y su aplicación en el proceso de búsqueda. Se introdujo la utilización de ontologías para la visualización y como su taxonomía representa un mecanismo para explotar la información semántica subyacente.

Si bien las ontologías juegan un rol principal en la creación de la Web Semántica, se espera que su utilización, en este caso en la visualización, sirva como herramienta importante para la búsqueda, navegación y representación de la información.

Como trabajo futuro se plantea la incorporación de otros servicios que requieran de elementos visuales, como por ejemplo un sistema de recomendación de documentos. Además se deja la posibilidad de extender la ontología de visualización con nuevos conceptos que permitan la descripción de servicios de visualización con mayor nivel de detalle o complejidad.

REFERENCIAS

- Baeza-Yates, R. y Ribiero-Nieto, B. (1999), *Modern Information Retrieval*, ACM Press, Addison-Wesley NY.
- Berners-Lee, T., Hendler, J. y Lassila, O. (2001), 'The semantic web', *Scientific American*.
- Castells, P. (2004), Neptuno: tecnologías de la web semántica para una hemeroteca digital. Article Llarg. Interacción 2004, Lleida. Maig 2004.
- Chen, C. (2002), Information visualization versus the semantic web, in 'Proceedings of Visualising the Semantic Web - VSW'.
- Consortium, W. W. W. (2004a), 'Resource description framework (rdf)'.
*En linea: <<http://www.w3.org/RDF/>> C02/04
- Consortium, W. W. W. (2004b), 'Semantic web'.
*En linea: <<http://www.w3.org/2001/sw/>> C02/04
- Consortium, W. W. W. (2004c), 'Web ontology language (owl)'.
*En linea: <<http://www.w3.org/OWL/>> C02/04
- Contreras, J. (2003), 'Duontology: an approach to semantic portals based on a domain and visualization ontology', *KTWeb*.
- Contreras, J. (2004), A semantic portal for the international affairs sector, in '14th International Conference on Knowledge Engineering and Knowledge Management (EKAW'04)', Springer-Verlag, Lecture Notes in Computer Science (LNCS) 3257:203-215.
- Fluit, C., Sabou, M. y van Harmelen, F. (2003), 'Supporting user tasks through visualization of light-weight ontologies', *Handbook on Ontologies in Information Systems*. Springer-Verlag.
- Fluit, C., Sabou, M. y van Harmelen, F. (n.d.), *Ontology-based Information Visualisation: Towards Semantic Web Applications. Visualising the Semantic Web*, 2nd edn, Springer Verlag.
- Fluit, C. y Wester, J. (2002), Using visualization for information management tasks, in 'Proceedings of the Sixth International Conference on Information Visualisation, IEEE', pp. 447-454.
- Gruber, T. (1995), 'Towards principles for the design of ontologies used for knowledge sharing', *International Journal of Human-Computer Studies* 43(5/6), 907-928.
- Guha, R., McCool, R. y Miller, E. (2003), Semantic search, in '12th International World Wide Web Conference (WWW2003)', Budapest, Hungary', pp. 700-709.
- Guzmán, J., Sánchez, A. y D., T. (2005), Desarrollo de un sistema de recuperación de información en JADE, in 'Congreso CICOM '05'.
- Hyvonen, E., Saarela, S. y Viljanen, K. (2003), Ontogator: Combining view- and ontology-based search with semantic browsing, in 'Proceedings of XML Finland 2003. Open Standards, XML and the Public Sector', Kuopio.
- Iglesias, C. (1996), A methodological proposal for multilingual systems development extending commonkads, in 'Proceedings of the 10th Banff Knowledge Acquisition for Knowledge-Based Systems Workshop', Vol. 1, Gaines, B. Musen, M.
- KARTOO (2004), 'Kartoo, versión 4'.
*En linea: <<http://www.kartoo.com>> C02/04
- MIII Groxis, I. (2004), 'Grokker, versión 2.1'.
*En linea: <<http://www.grokker.com>> C02/04
- Shah, U. (2003), Information retrieval on the semantic web, in '10th International Conference on Information and Knowledge Management', ACM Press.
- Ziegler, J., Kunz, C. y Botsch, V. (2002), Matrix browser - visualizing and exploring large networked information spaces, in 'International Conference on Computer Human Interaction SIGCHI 2002, Minneapolis', ACM Press.

