



Federación de servicios de búsqueda mediante un sistema semántico multiagente

Search Service Federation by means of a Multiagent Semantic System

◆ D. F. Barrero, O. García Población y D. R. López

Resumen

Las organizaciones suelen almacenar la información en distintos soportes, y por diversas consideraciones conviene articular mecanismos que permitan un acceso integrado a dicha información. Esta cuestión presenta una especial complejidad en el caso de que la información deba ser compartida entre distintas organizaciones. A continuación describimos un sistema que aborda parcialmente esta problemática, proporcionando integración de la búsqueda de recursos de una forma flexible, distribuida y compatible con la web semántica, que se adapta particularmente bien a la interoperación entre organizaciones.

Palabras clave: Recuperación de información, federación de servicios, servicios web, web semántica, middleware.

Summary

Organizations quite often store information in different supports, for several reasons, providing mechanisms to give an integrated access to this information is a recommended practice. The complexity of this task grows when some organizations have to share their information. This paper describes a system that faces it up; the system provides resource search integration in a flexible and distributed way, using semantic web technologies focused on the information interoperation between organizations.

Keywords: Information retrieval, service federation, web services, semantic web, middleware.

1.- Introducción

La evolución de las Tecnologías de la Información, en su intento de facilitar soluciones, ha creado diversos soportes para almacenar y gestionar la información. La aparición de las redes de telecomunicaciones y la informática personal a costes reducidos ha potenciado la producción de grandes volúmenes de información. Si a estos dos hechos le unimos que en muchas ocasiones los recursos de información están asociados a una escasa metainformación o sistema de gestión, podemos concluir que la gestión efectiva de toda la información es un problema de considerables dimensiones.

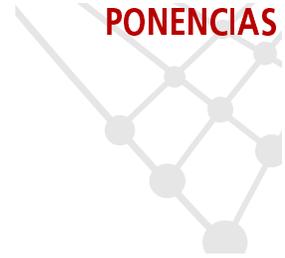
Gestionar la información dentro de una organización es de por sí una tarea compleja, pero si le unimos la frecuente necesidad de que distintas organizaciones, tecnológicamente independientes, interoperen compartiendo información, la tarea se vuelve todavía más compleja.

Presentamos un sistema, denominado Searchy, que permite integrar la búsqueda de recursos a través de sistemas de información heterogéneos, ubicados en unidades administrativas independientes, para lo cual ofrece una única interfaz de acceso a distintos sistemas de búsqueda.

A través de Searchy el usuario puede realizar una consulta sobre algún atributo del recurso que busca y el sistema de forma transparente accede a distintos sistemas de información, realiza la búsqueda y como resultado devuelve una descripción semántica de los recursos encontrados.

Searchy es fruto de la colaboración entre Rediris y la Universidad de Alcalá a través del programa PTYOC (<http://www.rediris.es/app/ptyoc>).

◆
El sistema, denominado Searchy, permite integrar la búsqueda de recursos a través de sistemas de información heterogéneos, ubicados en unidades administrativas independientes



2.- Objetivos

El objetivo principal de Searchy es proporcionar un mecanismo de búsqueda de recursos de información heterogéneos, para lo cual se crea una capa de abstracción de la consulta que permite formularla independientemente del formato local que adopte. En definitiva, se pretende realizar un sistema que federe distintos servicios de búsqueda, de manera que pueda accederse a todos ellos con una única interfaz y manejando un único formato de información.

Paralelamente al objetivo de federar distintos servicios de búsqueda, para que el sistema pueda tener una utilidad real, se consideran una serie de condiciones adicionales. La más restrictiva es que se debe poder acceder a sistemas de información que se encuentren en unidades administrativas distintas, y debe realizarse de manera no intrusiva, de manera que se garantice una completa autonomía en las decisiones tecnológicas.

Por otra parte, hay que garantizar la fácil extensibilidad del sistema, de manera que pueda acceder a nuevos sistemas de información con la máxima sencillez, incluyendo sistemas a priori desconocidos. Con el objeto de no sobrecargar al ya de por sí saturado personal técnico, el sistema debe ser lo más sencillo posible de administrar; de esta manera se plantea la necesidad de que el sistema sea completamente autónomo, lejos de complejos contenedores de aplicaciones, y que ofrezca una interfaz sencilla de manejar.

Para que la integración sea realmente efectiva hay que considerar dos aspectos de la misma: integrar procesos e integrar información, de manera que tanto la interfaz de acceso al servicio como el formato de la información que maneje el usuario sea uniforme, independientemente del soporte final al que se esté accediendo.

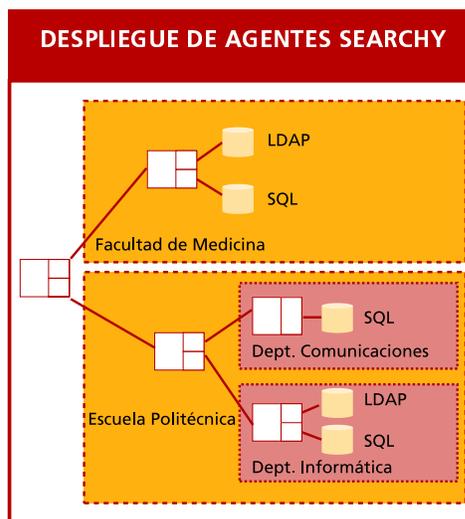
Se pretende realizar un sistema que federe distintos servicios de búsqueda, de manera que pueda accederse a todos ellos con una única interfaz y manejando un único formato de información

3.- Descripción de Searchy

Para cumplir los objetivos antes descritos se desarrolló Searchy. En esencia, Searchy es un metamotor de búsqueda distribuido, basado en el concepto de agente, que permite buscar recursos a través de sistemas heterogéneos utilizando descripciones semánticas.

El usuario que utiliza Searchy interactúa con un único sistema, que percibe como monolítico, cuando en realidad accede a una diversidad de sistemas distintos. Searchy se encarga de traducir la consulta desde el formato abstracto hasta los formatos locales manejados por cada sistema de información y, análogamente, traduce la respuesta desde el formato local a un formato abstracto, todo de manera transparente.

Desde la perspectiva del usuario, se envía una consulta sobre unos determinados campos y se obtiene como respuesta los distintos recursos que satisfacen la consulta, junto con una





El destinatario final de Searchy no es el usuario sino otras aplicaciones, típicamente una aplicación web que integrará el resultado de Searchy dentro de un sitio web

descripción semántica de dichos recursos. El uso que se haga de dichos resultados queda fuera del alcance de Searchy; en un uso típico el recurso se descargaría por otros medios utilizando la URL asociada, que se facilita en la descripción del recurso, pero cualquier otro uso es perfectamente válido.

En el corazón del sistema se encuentra el concepto de agente, que es la unidad básica que contiene toda la funcionalidad de Searchy y proporciona una importante serie de ventajas operativas. Los agentes son cooperativos, de manera que pueden colaborar entre ellos para completar la consulta del usuario, formando estructuras jerárquicas como la mostrada en la figura 1. En esta figura se puede observar una aplicación típica de Searchy compuesta de varios agentes que acceden a distintos sistemas de información.

En el ejemplo mostrado cada agente es responsable de acceder a una serie de soportes de información, que a su vez pueden ser otros agentes, por lo que puede adaptarse perfectamente a la división administrativa de una organización. Cada departamento puede compartir sus recursos de información proporcionando un agente público como punto de entrada. Así se consigue integrar la localización de recursos transversalmente a la organización (u organizaciones) y al mismo tiempo queda garantizada la independencia de la tecnología que utilicen internamente.

Junto al concepto de agente aparece otro concepto de importancia: el proveedor. La función de los proveedores es manipular las peculiaridades de cada sistema de información. Su funcionamiento es muy similar a los plug-in; introduciendo proveedores se amplía la funcionalidad del sistema, permitiéndole acceder a nuevos soportes de información.

El concepto de proveedor en este contexto es de gran potencia, su implementación hace del desarrollo de nuevos proveedores una tarea relativamente sencilla y además pueden incorporar algoritmos con una única condición: que acepten una consulta y devuelvan como resultado el valor de una serie de campos. Por lo tanto se puede abordar un amplio rango de problemas desde esta aproximación.

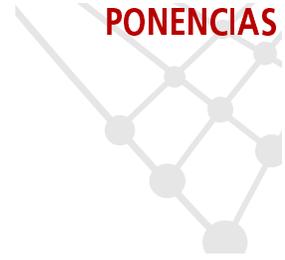
Conviene destacar que el destinatario final de Searchy no es el usuario, sino otras aplicaciones, típicamente una aplicación web que integrará el resultado de Searchy dentro de un sitio web. Por lo tanto Searchy no está concebido para ser utilizado por sí sólo, normalmente necesitará el desarrollo de un cliente específico para el entorno en el que se adopte.

Actualmente hay desarrollados tres proveedores que facilitan acceso a bases de datos SQL, servidores de directorio LDAP y al API de búsqueda de Google.

4.- La tecnología de Searchy

Como ya se comentó anteriormente, la integración de servicios de búsqueda tiene que hacer frente a dos problemas fundamentales: la integración de procesos y la integración de información. Ambos problemas han sido afrontados desde las últimas tecnologías disponibles en la Web.

Por una parte, cada agente expone su funcionalidad en forma de servicios web a través de SOAP (Simple Object Access Protocol). Con ello se proporciona una interfaz de acceso independiente del lenguaje y de la arquitectura que puede ser utilizado desde una gran cantidad de entornos, tanto web como no web. Además se obtienen otros beneficios, como es la posibilidad de utilizar WSDL (Web Services Description Language) y UDDI (Universal Description, Discovery and Integration) para el descubrimiento e invocación automática de agentes, entre un largo etcétera.



Para integrar la información se ha recurrido a la tecnología base de la web semántica: RDF (Resource Description Framework). A través de RDF se facilita la integración de Searchy en sistemas semánticos compatibles con la web semántica a la vez que puede utilizarse en entornos más tradicionales.

En la fase de desarrollo actual de Searchy no se soporta RDF en su totalidad, sino que está restringido a ser utilizado con un vocabulario específicamente creado para describir recursos. Dicho vocabulario se denomina Dublin Core y define un conjunto de quince términos que permiten describir recursos de la más diversa índole desde distintas perspectivas que abarcan conceptos como título, materia, autores, etc.

5.- Conclusiones

La aproximación que Searchy realiza a la integración de sistemas ha demostrado ser satisfactoria. Es un sistema de búsqueda altamente distribuido, escalable y modular que se amolda a un amplio espectro de situaciones distintas. La utilización de proveedores permite aumentar fácilmente su funcionalidad, soportando nuevos sistemas de información y le proporciona una gran capacidad para adaptarse a nuevas aplicaciones desconocidas a priori.

Searchy está disponible libremente a través de la licencia GPL y se encuentra en pleno proceso de desarrollo con distintas líneas de trabajo. Por una parte se plantea la conveniencia de dar soporte pleno a RDF, en cuyo caso el sistema trabajaría no con descripciones de recursos, sino directamente con la traducción semántica de los mismos, ampliando de forma espectacular su rango de aplicaciones; por otra parte hay una importante cantidad de posibles proveedores que pueden elaborarse para dar soporte a servidores HTTP y FTP, ficheros en formato Word u OpenOffice, o en otra línea, proveedores que localicen sistemas de búsqueda automáticamente mediante WSDL/UDDI o OWL-S.

En la fase de desarrollo actual de Searchy no se soporta RDF en su totalidad

Bibliografía

- Página web del proyecto Searchy. Consultado en: <http://sourceforge.net/projects/jsearchy>. Fecha de acceso: 19 de octubre de 2004.
- Colaboración de RedIRIS en la realización de trabajos académicos. Consultado en: <http://www.rediris.es/app/PTYOC/>. Fecha de acceso: 19 de octubre de 2004.
- BARRERO, DAVID F.; LÓPEZ, DIEGO R.; GARCÍA POBLACIÓN, ÓSCAR. "Distributed Metainformation Searching: A New Approach to Information Retrieval in the Age of the Semantic Web". Presentado en TERENA Networking Conference. 2004. Rodas, Grecia.
- BARRERO, DAVID F.; CRIADO, J. IGNACIO. "Integrando la Información de las Administraciones Públicas en la Gestión del Conocimiento. Una Solución desde la Web Semántica y los Servicios Web". Presentado en VIII Jornadas TECNIMAP. 2004. Murcia, España.
- BERNERS-LEE, TIM; HENDLER, JAMES; LASSILA, ORA. 2001. The Semantic Web, Scientific American 284 (5):28-31.
- DCMI. DCMI Metadata Terms. Consultado en: <http://www.dublincore.org/documents/dcmi-terms>. Fecha de acceso: 27 de junio de 2004.
- HENDLER, JAMES. 2001. Agents and the Semantic Web. IEEE Intelligent Systems Journal, 16(2):30-37, Marzo-Abril.



- MILLER, ERIC; MANOLA, FRANK. 2004. (eds.). RDF Primer, <http://www.w3.org/TR/rdf-primer/>. Fecha de acceso: 19 de octubre de 2004
- MITRA, NILO, 2003. (ed.). SOAP Version 1.2 Part 0: Primer. <http://www.w3.org/TR/soap12-part0/>. Fecha de acceso: 19 de octubre de 2004.
- Programa PTYOC. Fecha de acceso 19 de octubre de 2004. <http://www.rediris.es/app/ptyoc>

David. F. Barrero
(dfbarrero@wanadoo.es)
Óscar García Población
(oscar@aut.uah.es)
Dpto. de Automática
Universidad de Alcalá
Diego R. López
(diego.lopez@rediris.es)
Área de Middleware
RedIRIS