

Flexible Management of Distributed Contents

◆ L. Sánchez Fernández, J. Villamor, J. Arias et al.

Resumen

El desarrollo de las tecnologías de la información y de las comunicaciones y, especialmente, el desarrollo de Internet ha hecho que, en la actualidad, existan grandes volúmenes de información disponibles en formato electrónico. En diferentes campos ha surgido la necesidad de gestionar dicha información, que anteriormente se archivaba en formato papel. Para llevar a cabo esta tarea, han aparecido varios sistemas de gestión de contenidos, casi todos orientados a la publicación de contenidos en un sitio web centralizado. No soportan el cooperar con un conjunto de sistemas de gestión de contenidos distribuidos en red. El presente artículo presenta una propuesta que pretende diseñar e implementar una arquitectura para soportar la gestión eficiente y flexible de contenidos distribuidos.

Palabras clave: Gestión de contenidos, Servicios Web, Web Semántica.

Summary

The development of information and communication technologies and the expansion of the Internet means that nowadays there are huge amounts of information available via these emergent media. The need to manage such information, which was in the past stored on paper media, has become apparent in different fields. A number of content management systems have appeared with the aim of achieving this task. Most of these systems are oriented towards Web publishing on a central site, and they do not support collaboration among several, distributed sources of managed content. In this paper we present a proposal for an architecture for the efficient and flexible management of distributed content.

Keywords: Content management, Web Services, Semantic Web.

◆
En diferentes dominios de aplicación existe la necesidad de poder ofrecer un servicio que permita gestionar, producir y ofrecer información de calidad

1.- Introducción

En diferentes dominios de aplicación existe la necesidad de poder ofrecer un servicio que permita gestionar, producir y ofrecer información de calidad. Las entidades que se dedican a la producción y/o archivo de contenidos deben enfrentarse con una serie de problemas a resolver:

- **Gestión del proceso de creación y modificación de contenidos.** Este trabajo requiere de un sistema que permita que las diferentes personas que trabajan no se interfieran, de forma que los cambios que introduzca una no eliminen el trabajo realizado por otra. Es decir, se necesita un mecanismo de control de versiones que permita el trabajo independiente pero coordinado de las diferentes personas.
- **Gestión de perfiles de usuario.** En un entorno en el que se crea y accede a contenidos, deben definirse mecanismos para identificar quién puede crear, modificar o consultar documentos. También puede ser necesario arbitrar mecanismos de revisión de documentos que han sido creados, como paso previo a su difusión interna o pública.
- **Necesidad de arbitrar algún mecanismo de decisión sobre qué contenidos de los almacenados en el sistema se deben publicar.** Esta decisión ha de tener en cuenta dos tipos de factores: técnicos (en el caso de que los contenidos se publiquen en un sitio web: validar si los enlaces están bien, cuál es el lugar adecuado para colocar un determinado contenido, etc.) y políticos (confianza en la información, confidencialidad, etc.).
- **Problemas derivados de la integración de varios fragmentos producidos** (por ejemplo, para producir una página web).
- **Validación.** Los contenidos y funcionalidades ofertadas por el sistema deben probarse para detectar errores, inconsistencias (enlaces que apuntan a sitios inexistentes, scripts que no funcionan, etc.).



Los Servicios Web son una tecnología para facilitar la cooperación de aplicaciones remotas que ha sido impulsada por un conjunto de empresas y está actualmente en proceso de estandarización por el World Wide Web Consortium

En este artículo presentamos una propuesta de una arquitectura (actualmente en proceso de desarrollo) para un sistema de gestión de contenidos flexible y distribuido: el sistema Infoflex. Este sistema ofrece las facilidades requeridas habitualmente para un sistema de gestión de contenidos. Dichas facilidades se implementan utilizando las siguientes tecnologías:

- Uso de XML para la descripción (y en el caso del contenido textual su encapsulación) de los contenidos almacenados. Los documentos XML se almacenarán en una base de datos XML para facilitar su recuperación. Asimismo, el uso de XML permitirá exportar los contenidos a diferentes formatos (utilizando hojas de estilo XSLT).
- Un sistema de workflow sencillo con las siguientes funcionalidades:
 - definición de procesos con interfaz gráfica, utilizando un lenguaje estándar basado en XML
 - definición de roles y actores para control de acceso
 - interfaz de interacción con actores
 - interfaz gráfica sencilla de gestión
 - enrutado de procesos
 - planificación de tareas
- Un sistema de gestión de versiones, desarrollado sobre la herramienta de libre distribución CVS.

Adicionalmente, el sistema permitirá la consulta y recuperación de contenidos de forma remota. Se pretende desarrollar un sistema que permita a un usuario (humano o sistema informático) localizar (incluso sin conocer previamente su ubicación) diferentes sistemas de gestión de contenidos distribuidos y heterogéneos y realizar una única consulta sobre varios de ellos.

Para conseguir este objetivo se plantea utilizar tecnologías de Servicios Web y Web Semántica [1]. Los principios básicos se presentan a continuación.

2.- Los Servicios Web y su aplicación en Infoflex

Los Servicios Web [3] (Web Services) son una tecnología para facilitar la cooperación de aplicaciones remotas que ha sido impulsada por un conjunto de empresas (IBM, HP, SUN, Microsoft, ...) y está actualmente en proceso de estandarización por el World Wide Web Consortium (W3C). Esta tecnología pretende ser utilizada para la cooperación de aplicaciones inter-empresa, aunque también podría aplicarse al caso de aplicaciones intra-empresa.

La arquitectura de los Servicios Web se basa en dos principios básicos:

- Uso de XML, como formato de intercambio de datos y descripción de servicios.
- Uso de protocolos estándar de Internet y universalmente utilizados para el intercambio de mensajes, tales como HTTP y SMTP (el modelo de Servicios Web no restringe el intercambio de mensajes a estos dos protocolos, pero son ejemplos típicos de protocolos que pudieran ser usados).


El modelo de interoperabilidad propuesto por la tecnología de Servicios Web encaja perfectamente con los objetivos que nos proponemos resolver para el acceso a contenidos distribuidos en este proyecto.

Recordemos que el objetivo es permitir el que un usuario, humano o aplicación software, pueda lanzar una búsqueda de información sobre gestores de contenidos de cierto dominio de aplicación (agencias de noticias, portales B2B de cierto ámbito, etc.). Se supone que no se conoce inicialmente

las entidades en las que buscar (al menos no todas), por lo que se debe habilitar un mecanismo para descubrir entidades que proporcionen contenidos en el ámbito de interés. Los Servicios Web proporcionan un mecanismo de publicación y descubrimiento que resuelve perfectamente este objetivo.

Del mismo modo, como los proveedores de contenidos pueden ser empresas diferentes, que utilicen diferentes plataformas, entornos, etc., –tal y como se ha explicado anteriormente–, los Servicios Web están especialmente indicados en este tipo de situaciones.

Finalmente, el hecho de que los sistemas de gestión de contenidos que se plantean utilizan documentos XML para la descripción del contenido almacenado implica que las búsquedas se harán contra documentos XML, por lo que el hecho de que entre el cliente y el servidor se intercambien documentos XML (como ocurre con los Servicios Web) simplifica el procesamiento de las consultas.


Ahora que la producción digital de textos es casi universal, se trata de construir una web, sobre que los recursos puedan ser interpretados por las máquinas más que ser visualizados por el hombre

3.- La Web Semántica y su aplicación a Inflex

Ahora que la producción digital de textos es casi universal, se trata de construir una web, sobre que los recursos puedan ser interpretados por las máquinas (por agentes software), que para que puedan ser visualizados por el hombre. Para ello, en el año 2001 el W3C lanzó el Semantic Web Activity Group [4] que trabaja en dos ámbitos:

- Metadatos: se desarrollan mecanismos para ser descritos por máquinas y de manera legible contenidos representados en formatos de publicación web (HTML, XML). Estos mecanismos se instrumentan utilizando, entre otros, el lenguaje estándar del W3C RDF (Resource Description Framework) [5].
- Ontologías: por encima del nivel de descripción se define un nivel de interpretación basado en la utilización de ontologías, que son descripciones de un dominio concreto de conocimiento (conceptos y relaciones entre ellos). Estas ontologías se representarán utilizando el lenguaje en proceso de estandarización por el W3C OWL (Ontology Web Language) [6].

En el marco del sistema Inflex, para resolver el problema de las consultas sobre proveedores de contenidos que utilizan bases de datos heterogéneas, se propone utilizar una ontología, definida con OWL, que describa los conceptos y sus relaciones y atributos en el dominio de aplicación de interés. El uso de ontologías para tratar con la heterogeneidad de varias fuentes de datos ha sido utilizado previamente, por ejemplo en [2].

Dado que esta ontología es general, nos planteamos ubicarla dentro de un servicio global en el que deberán registrarse los gestores de contenidos. A este servicio le llamaremos “Servidor de Consultas” que es a la vez un Servicio Web.

El proveedor de contenidos que desee ofrecerlos deberá registrarse en dicho servicio y proporcionar información sobre su modelo de datos: parte de los conceptos soportados por la ontología general que contemplan sus contenidos y relación entre el modelo de datos del proveedor de contenidos y la ontología general.

Por lo tanto, los proveedores de contenidos deberán realizar un doble registro: primero en un directorio para ofrecer su interfaz a posibles clientes y segundo en el Servidor de Consultas con el fin de permitir las consultas semánticas. Cuando un cliente (humano o aplicación) desee hacer una consulta seguirá los siguientes pasos:

- 1) Buscar en el directorio Servicios Web de acceso a proveedores de contenidos.



Aunque el sistema propuesto está en el inicio de su desarrollo, confiamos en que un sistema como éste pueda ser de utilidad en entornos como portales B2B, agencias de noticias o portales de sindicación de contenidos

2) Seleccionar uno o varios proveedores de contenidos. Por cada uno de ellos, consultar al Servidor de Consultas, indicando la consulta a realizar en algún lenguaje XML (por definir). El Servidor de Consultas puede dar dos respuestas:

- El modelo de datos del proveedor indicado no es compatible con la consulta solicitada.
- El modelo de datos del proveedor indicado es compatible con la consulta solicitada, en cuyo caso se devuelve equivalente en el modelo de datos del proveedor de contenidos.

3) Acceder a los proveedores de contenidos escogidos para realizar la consulta y obteniendo los resultados compatibles con ella.

4.- Conclusiones

En esta ponencia hemos presentado la propuesta de una arquitectura que estamos actualmente desarrollando para la gestión de contenidos flexible y distribuida basada en el uso de tecnologías web avanzadas, tales como XML, Web Semántica y Servicios Web.

Aunque el sistema propuesto está todavía en el inicio de su desarrollo, y no podemos presentar resultados sobre pruebas del mismo, confiamos en que un sistema como éste pueda ser de utilidad en entornos como portales B2B, agencias de noticias o portales de sindicación de contenidos.

Referencias

- [1] Berners-Lee, T.; Hendler, J.; Lassila, O., "The Semantic Web". "Scientific American". Vol. 284, nº 5. 2001. 34-43.
- [2] Brisaboa, N.; Penabad, M.; Places, A.; Rodríguez, F., "Ontologies for Database Federation". "UPGRADE". Vol. III, n. 3. 2002. 52-61.
- [3] World Wide Web Consortium. Web Services Activity. Consultado en: <http://www.w3.org/2002/ws/> 28-10-2003.
- [4] World Wide Web Consortium. Semantic Web Activity. Consultado en: <http://www.w3.org/2001/sw/> 28-10-2003.
- [5] World Wide Web Consortium. Resource Description Framework (RDF). Consultado en: <http://www.w3.org/RDF/> 28-10-2003.
- [6] World Wide Web Consortium. OWL Web Ontology Language Reference. Consultado en: <http://www.w3.org/TR/owl-ref/> 28-10-2003.

Este trabajo ha sido parcialmente financiado por el MCyT a través del proyecto TIC2003-07208 "InfoFlex"

Luis Sánchez Fernández, Jesús Villamor Lugo
(luis@it.uc3m.es), (jvl@it.uc3m.es)

Jesús Arias Fisteus, Norberto Fernández García
(jaf@it.uc3m.es), (berto@it.uc3m.es)
Dpto. de Ingeniería Telemática,

Tomás Nogales Flores, Antonio Hernández Pérez
(nogales@bib.uc3m.es), (tony@bib.uc3m.es)

David Rodríguez Mateos
(pirio@bib.uc3m.es)

Dpto. de Biblioteconomía y Documentación - UC3M