

Aplicación de los sistemas de tratamiento de bibliotecas digitales al Sistema de Información del Patrimonio Histórico Andaluz

M. J. Escalona
J. Torres
M. Mejías

*Dpto. Lenguajes y Sistemas Informáticos
Universidad de Sevilla*

Resumen

El desarrollo de un sistema de información de difusión del patrimonio es una tarea compleja, debido a los diferentes aspectos que se deben tener en cuenta en su diseño. Por un lado, la información se encuentra almacenada en diferentes medios: imágenes, cadenas de texto, documentos, etc. En este sentido, es necesario el uso de herramientas multimedia, y más concretamente hipermedia, para su difusión por la red. Por otro lado, los altos requisitos de almacenamiento de información y de protección, así como su consulta hacen que los sistemas de información del patrimonio histórico presenten complejos requisitos funcionales. Debido a la variedad de necesidades, en este documento vamos a presentar las bibliotecas digitales como una solución al problema real que se plantea en la gestión del patrimonio histórico dentro del Instituto Andaluz de Patrimonio Histórico. Para ello, proponemos una arquitectura basada en bibliotecas digitales, que permita el acceso uniforme a bases de datos heterogéneas y distribuidas, utilizando para el intercambio de información el estándar XML.

Introducción

El avance de las tecnologías y de las comunicaciones está haciendo que cada día los usuarios tengan más información a su alcance. En este sentido, el Instituto Andaluz de Patrimonio Histórico (I.A.P.H.) está desarrollando el Sistema de Información de Patrimonio Histórico Andaluz (SIPHA) para poder difundir la información que recopilan y con la que catalogan el patrimonio histórico. Sin embargo, en la actualidad, el SIPHA no es un sistema sencillo puesto que se caracteriza por tener una serie de requisitos que lo transforma en un sistema complejo y de difícil desarrollo. El uso de las bibliotecas digitales es una solución adecuada para este tipo de sistemas.

El principal problema que se presenta es la dispersión de las fuentes de información y la no uniformidad de criterios en la descripción de la estructura de la misma en función de la catalogación de los bienes. La información se aporta desde diferentes puntos geográficos y en diferentes soportes, además de en diferentes formatos de almacenamiento.

Dada la complejidad de los sistemas que pueden hacer frente a esta situación, en los cuales se deben gestionar múltiples y variadas bases de datos, es necesario definir un entorno arquitectónico para los mismos que permita la definición de sistemas distribuidos, abiertos y adaptados al uso de los más modernos entornos tecnológicos, como es el caso de internet (Brisaboa, 1999). Para facilitar la inclusión de estos sistemas en las redes internet es necesario el uso de herramientas y lenguajes de descripción normalizados, como por ejemplo XML, que faciliten la integración de los diferentes elementos que en los mismos intervienen.

Planteamiento del problema

En este apartado vamos a describir el problema que vamos a abordar a lo largo del documento, detallando sus módulos y problemas fundamentales.

Vamos a plantearnos el desarrollo de un sistema de información del patrimonio histórico de Andalucía (SIPHA) que se mostrara desde el servidor del IAPH y que a él pudiesen acceder de manera concurrente tanto los catalogadores de cada equipo de trabajo, que por supuesto pueden encontrarse en diferentes puntos geográficos (normalmente en Andalucía), como las personas que simplemente quieren realizar consultas (este tipo de usuario genérico puede acceder, en principio, desde cualquier parte del mundo) como los propios profesionales del IAPH.

Vamos a centrarnos en la parte de este sistema que controlaría el Patrimonio Inmueble Andaluz. En este sistema de información de patrimonio histórico se va a almacenar y gestionar toda la información referente a los bienes patrimoniales inmuebles. Cada bien puede ser catalogado desde tres puntos de vista: arquitectónico, arqueológico o etnológico. A esto es lo que se denomina caracterización del bien. Así, si un bien tiene interés arqueológico y arquitectónico, tendrá ambos valores en su caracterización.

En el IAPH existe un sistema de información sectorial por cada una de las ramas: sistema de información de patrimonio arquitectónico (SIBIA), sistema de información de patrimonio arqueológico (ARQUEOS) y sistema de información de patrimonio etnológico (ETNO). La información que en ellos se almacena podemos agruparla por módulos: módulo de identificación y localización, módulo de descripción, módulo de análisis, el módulo de protección, etc. Parte de los datos que se incluyen en estos módulos están normalizados, pudiendo tener sólo los valores que se encuentren en un almacén de términos que se denomina Tesauro.

Por otro lado, las tres bases de datos (Arqueológica, Arquitectónica y Etnológica) tienen otros módulos de información denominados de información documental y que se dividen a su vez en tres: información bibliográfica (que contiene los datos referentes a los libros que hablan de un bien determinado), información documental (que guarda la referencia a los documentos que tienen que ver con el bien) e información gráfica (que almacena las fotos digitalizadas del bien). Estos módulos documentales se nutren de la información que se toma de lo que denominaremos bases de datos documentales. Las bases de datos documentales, como se muestra en la figura 1, se dividen en tres: bibliográfica, documental y gráfica.

Aunque los tres sistemas de información tienen los mismos módulos, hay datos que no interesan a todos. Por ejemplo, el autor de un bien tiene interés para los arquitectos, pero los arqueólogos, en la mayoría de los casos, no lo conocen ni lo necesitan. Sin embargo, datos como el estado de protección en el que se encuentra, o por los que ha pasado, un monumento o sus datos de localización son comunes para las tres ramas.

Si pensamos en un sistema diferente para cada área que dé soporte a cada una de ellas, cuando hay bienes que son catalogados por más de un área temáti-

ca, la información común se duplica, teniéndose así los clásicos problemas que causa la redundancia. Pero además, el hecho de tener la información dispersa hace que su difusión sea compleja. Así, si un usuario necesitara los monumentos de estilo neoclásico, necesitaría buscar en los tres sistemas, pudiendo incluso recibir el mismo monumento por tres vías diferentes; resulta pues muy interesante la integración de todos ellos.

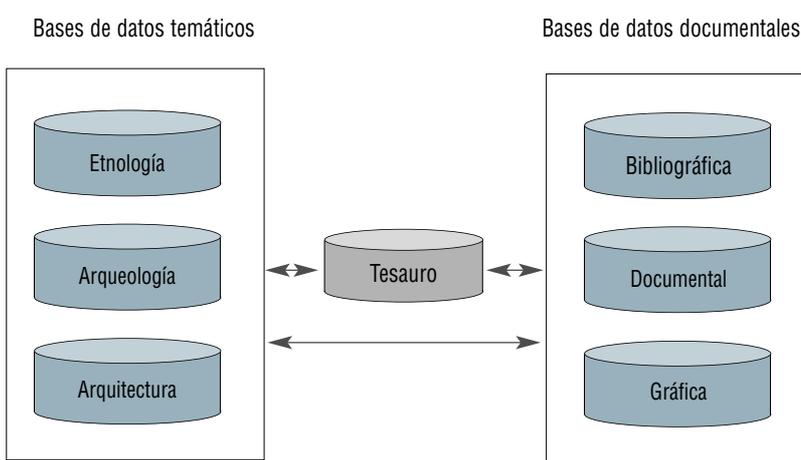
Por otro lado, resulta conveniente poder definir vistas independientes dentro de estos tres sistemas. Por ejemplo, que en un momento dado, un etnólogo pudiese obviar la información arqueológica y la arquitectónica.

Otro punto importante de este sistema es el Tesauro de términos que lleva asociado. Existen una serie de campos y valores que están normalizados dentro de la terminología que controla el Tesauro del IAPH. Así, por ejemplo, la tipología, el estilo o las actividades para las que se destina (o se destinaba) el bien, sólo pueden tomar valores concretos del Tesauro.

También habría que indicar que en este sistema encontramos información de distinta naturaleza. Por un lado, encontramos información que podría almacenarse en una base de datos relacional clásica: datos de tipo cadena de textos o numéricos. Por otro lado, encontramos información como los libros, los documentos o los expedientes de protección o conservación, las publicaciones sobre el estado de protección que se hacen en el BOE o en el BOJA sobre el bien, etc., para los que sería más aconsejable una base de datos documental. Y por último, si añadimos la información gráfica observamos que ninguno de estos sistemas de almacenamiento nos dan el soporte necesario para nuestra aplicación.

Tampoco nos podemos olvidar de los tipos de usuarios a los que va dirigida la información y los sistemas de consulta de los que disponen. Así, los catalogadores pueden dar de alta y modificar la información del sistema que les corresponde a su entorno de trabajo y los usuarios solamente pueden realizar consultas. Con respecto a las consultas, son complejas y variadas. Así, puede haber consultas tan sencillas como listar todos los monumentos de una ciudad y consultas en las que, por ejemplo, se intenten localizar todas las fotos de edificios, sean del estilo que sean, en las que se encuentre una cruz.

Figura 1
Bases de datos componentes del SIPHA de Bienes Inmuebles



Justificación del uso de bibliotecas digitales

El SIPHA, como podemos concluir del apartado anterior, tiene una serie de características que lo hace diferente de un sistema de información tradicional. En él se mezclan las necesidades de integración de elementos multimedia, con la posibilidad de navegar por la aplicación y con importantes requisitos de almacenamiento, funcionalidad, concurrencia y seguridad. Un SIPH es pues un híbrido entre lo que constituye una aplicación de gestión clásica, debido a

esos requisitos de almacenamiento, funcionalidad, concurrencia y seguridad; y entre lo que se conoce como aplicación hipermedia, aquella que gestiona información de diferente naturaleza: imagen, sonido, animación, etc., por la que se puede navegar mediante hiperenlaces (Martínez, 1997). En la figura 2 encontramos esquematizada esta idea.

Una biblioteca digital es una biblioteca que ha sido extendida y mejorada mediante la aplicación de la tecnología digital. Es la unión de ordenadores, sistemas de almacenamiento y redes de comunicaciones con el contenido y el software necesario para reproducir, emular y extender los servicios proporcionados por las bibliotecas convencionales. Una biblioteca digital debe cumplir todas las tareas de una biblioteca convencional y explotar las ventajas de la tecnología digital en el almacenamiento, la búsqueda y las comunicaciones, además de la integración de nuevos tipos de medios: textos, imágenes, sonidos, vídeos, animaciones, etc.

La biblioteca digital proporciona a una comunidad de usuarios un acceso coherente a repositorios de información grandes y organizados. Las bibliotecas digitales son construidas, recogiendo y organizando la información, por una comunidad de usuarios y sus funcionalidades son acordes a las necesidades de información de dicha comunidad. Las posibilidades de los usuarios para acceder, reorganizar y utilizar estos repositorios están enriquecidas con las capacidades de la tecnología digital.

Entre las ventajas de las bibliotecas digitales sobre las tradicionales se puede citar la posibilidad de contener documentos cambiantes y transitorios, además de los documentos fijos y permanentes propios de las bibliotecas tradicionales. También es una ventaja el hecho de permitir el uso de la información de forma cooperativa y no aislada.

El campo de las bibliotecas digitales es la unión de varios subcampos de otros dominios, y abarca muchos temas de interés. Es un tema que actualmente se está desarrollando y para el que existen publicaciones específicas (D-Lib magazine, Int. Journal on Digital Libraries de Springer), así como informes especiales en publicaciones más genéricas (Proc. of the ACM, ACM Multimedia, IEEE Multimedia). También existen congresos especializados en el tema, como la ACM Conference on Digital Libraries y el IEEE Advances in Digital Libraries (ADL).

Evidentemente, entre las áreas de aplicación más importantes, se encuentran el Turismo, la Educación, la Investigación, la Cultura y el Ocio. Como proyectos concretos se pueden citar las Bibliotecas nacionales, los Museos, los Archivos Históricos (como el Archivo General de Indias de Sevilla, uno de los pioneros), las redes comunitarias de servicios (poblados electrónicos), los repositorios de Technical Reports (sobre todo en Computer Science), datos geográficos (Medio Ambiente) o de Literatura, y las publicaciones electrónicas (como la ACM Digital Library).

Según todo esto podemos concluir que las bibliotecas digitales se convierten en una herramienta adecuada para el desarrollo del sistema de información de SIPHA.

Propuesta de arquitectura

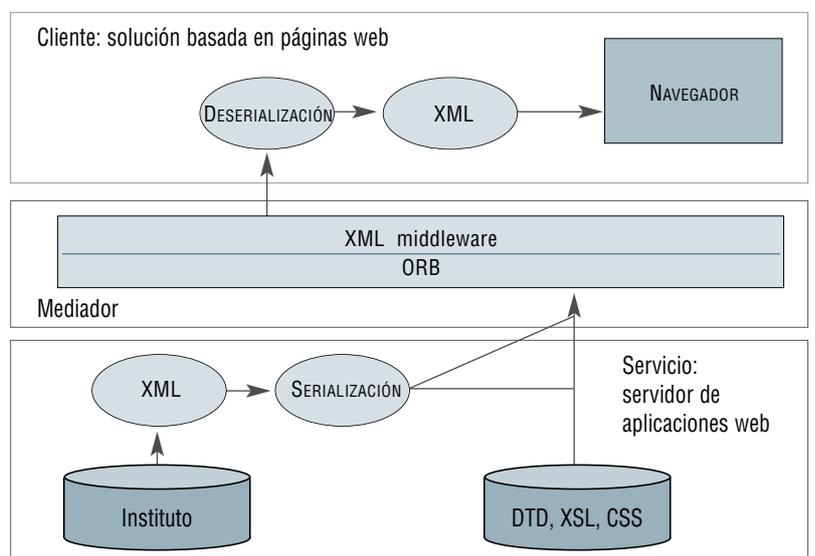
La propuesta de arquitectura que desde el entorno de las bibliotecas digitales se puede realizar se basa en el esquema adjunto:

Esta arquitectura es lo que se conoce como una arquitectura basada en cliente-servidor. En ella, las bases de datos del IAPH seguirían teniendo su estructura inicial y se encontrarían ubicadas en el Instituto (es lo que en la figura se recoge como "instituto"). La idea no consiste en reestructurar todas las bases de datos sectoriales, la idea propuesta desde el entorno de las bibliotecas digitales se basa en colocar una serie de capas que hagan al usuario tener una visión global e integrada de toda la información que sobre los bienes inmuebles se tiene. Para ello, se propone el uso de XML (eXtensible Markup Language). XML es un mecanismo que permite la distribución de la información entre el nivel del mediador y el del servidor, así como el del mediador y el cliente, y además está normalizado. XML, ha sido propuesto

Figura 2
Relación entre los sistemas de tratamiento de bibliotecas digitales, las aplicaciones clásicas y las aplicaciones multimedia



Figura 3
Propuesta de arquitectura para el IAPH



por World Wide Web Consortium (WWW, 2000) y está destinado al intercambio de contenido o datos específicos entre aplicaciones y se suele definir como un lenguaje de inclusión de metadatos. Utilizando XML definimos las características de los medios digitales presentes en la respuesta a una consulta: tipo, tamaño, localización, enlaces, etc.

Una vez que hemos determinado el tipo de información que se va a mostrar mediante XML, se deben definir lo que se conoce como DTD (Document Type Definition). Los DTD son documentos que complementan a XML y que albergan la descripción de los elementos que pueden aparecer en un archivo XML y la organización que van a tener en la salida.

Hasta este punto ya tendríamos definidos los datos que se pueden mostrar y los formatos que van a tener en la salida. Quedaría por estudiar cómo van a realizarse las presentaciones de las consultas. Para ello, las tecnologías más utilizadas son las hojas de estilo aplicadas tanto a fuentes HTML como a fuentes XML. Para el caso de XML se han definido los formatos CSS (Cascade Style Sheet) y XSL (eXtensible Stylesheet Language).

De esta forma, en el nivel del servidor (es decir, en el entorno del IAPH) se tendrían dos tipos de bases de datos: las denominadas "instituto", que son las bases de datos sectoriales descritas en la figura 1 y que mantendrían su estructura y que contendrían la información objeto de tratamiento, y por otra parte se encuentran las denominadas como "DTD-XSL-CCS", que incluyen los metadatos y las hojas de estilo, las cuales definen, respectivamente, los tipos de datos, su organización y la forma de presentarlos.

El nivel intermedio de esta propuesta de arquitectura lo constituye el mediador, que opera con el nivel inferior a través de un ORB (Object Request Broker) normalizado que independiza el acceso a las bases de datos del sistema de consulta. En este nivel también se encuentra el XML middleware. El XML middlewa-

re se encarga de recibir las peticiones realizadas desde el nivel cliente y de traducirlas de manera que el nivel servidor las entienda. En el camino inverso, es capaz de tomar los resultados por parte del servidor y de mostrarlo al cliente de forma homogénea.

Por último, en el nivel superior se encuentra el cliente, que dispone de un navegador que le presenta la interfaz de usuario a través de la cual puede solicitar información o modificarla, dependiendo de las prioridades de uso del sistema que tenga. La información que llega al navegador está en formato XML, conteniendo los enlaces a los documentos necesarios para la descripción de los datos (DTD) y su presentación al usuario (hojas de estilo).

Finalmente, decir que los procesos de serialización y deserialización que aparecen en los niveles servicio y cliente, respectivamente, tienen como fin facilitar el flujo de información entre capas.

Una vez realizada esta propuesta podemos hacer un análisis de sus ventajas. Por un lado, el usuario de la aplicación es totalmente ajeno a la complejidad interna que puede tener el sistema. Así, un usuario consultor verá igual de complejo el solicitar una lista de todos los yacimientos de Sevilla como el hacer una petición de todas las fotos de inmuebles en las que aparezca una cruz. Será el mediador el que se encargue de hacer que el formato de estas consultas sea el adecuado y vaya a la base de datos en concreto que pueda dar los resultados solicitados.

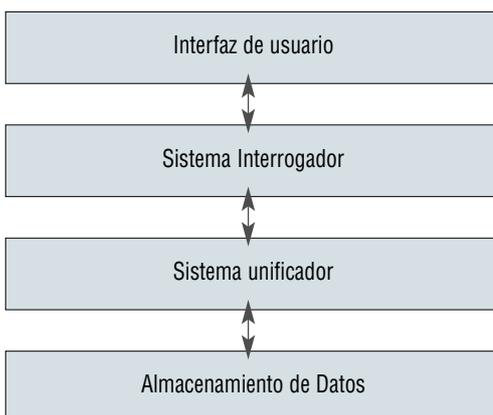
Por otro lado, el sistema resulta totalmente seguro puesto que el mediador, a través de claves es capaz de detectar los permisos que tiene un usuario. De esta forma, si un usuario es capaz de ofrecer la clave necesaria para dar de alta un bien, podrá modificar las bases de datos. Además, información que no debe salir del ámbito del Instituto, como las direcciones de los bienes no protegidos, el nombre de los propietarios, etc. serían restringidos indicando al DTD correspondiente que no debe mostrarse.

Para terminar, y a título informativo, hay que decir que la arquitectura que aquí se ha propuesto concreta una definición de arquitectura genérica basada en cuatro capas y que se muestra en la figura 4.

Estas capas se definen a continuación:

1. Interfaz de Usuario: Permite el acceso a la información a través de consultas que se envían al Sistema Interrogador y la visualización de los resultados de esas consultas.
2. Sistema Interrogador: Traslada las peticiones del usuario a un lenguaje de consulta de datos.
3. Sistema Unificador: Desglosa las peticiones del usuario para recoger la información de las bases de datos correspondientes.
4. Almacenamiento de Datos: Almacena la información de forma distribuida en múltiples y variadas bases de datos.

Figura 4
Arquitectura genérica



La aplicación de este tipo de arquitecturas pretende que la configuración del sistema construido sea lo más abierta posible. Así, el modelo es lo que se conoce como un modelo escalable, es decir, permite la incorporación de nuevas bases de datos y de nuevos medios digitales que se adapten a las nuevas tendencias tecnológicas.

Conclusiones

En este trabajo se ha planteado la problemática que rodea a la definición de sistemas de información de patrimonio histórico, sus similitudes con los sistemas informáticos que tratan con información multimedia mediante hiperenlaces y el adoptar como solución para la definición de los mismos el uso de sistemas de tratamiento de bibliotecas digitales

Se ha presentado una solución arquitectónica concreta a este problema, basada en la definición de arquitectura en cuatro capas de Witten y adaptada al uso de internet y XML como técnica normalizada de descripción de datos.

Aunque los sistemas de tratamiento de bibliotecas digitales se han propuesto como solución para los sistemas de información de patrimonio histórico, estos sistemas pueden constituir también una solución adecuada para otros entornos, en los que se integrarán necesidades de gestión de información multimedia y complejos requisitos funcionales, de seguridad, de eficiencia y eficacia.

Referencias

Brisaboa, 1999: Brisaboa, Nieves. M. J. Durán, C. Lalín, J. R. López, J. R. Paramá, M. R. Penabad, A.S. Places, *Arquitectura para Bibliotecas Virtuales*, 1999.

Lesk, 1997: Lesk, Michael. *Practical Libraries: Books, Bytes & Bucks*. Morgan Kaufmann. 1997.

Martínez, 1997: J. M. Martínez, J. R. Hilera, J. Martínez, J. A. Gutiérrez, *Orientación a Objetos en la Documentación Hipermedia*, Universidad de Alcalá de Henares, Departamento de Ciencias de la Computación.

Witten, 1996: Witten, I. Craig Nevill-Manning, Sally Jo Cunningham. Building a Digital Library for Computer Science Research: Technical Issues. *Proceedings of the 19th Australasian Computer Science Conference*. 1996.

WWW, 2000: World Wide Web Consortium.
www.w3c.org, 2000.