

Aplicación Integrada de la Biblioteca Digital del Patrimonio Histórico Andaluz

M^aJosé Escalona, Manuel Mejías, Jesús Torres, Juan M. Cordero, Mariano G. Romano

Dpto. Lenguajes y Sistemas Informáticos
Universidad de Sevilla
escalona@lsi.us.es

Resumen. En este trabajo se describen las necesidades del Instituto Andaluz de Patrimonio Histórico para desarrollar el Sistema de Información Integrado de Patrimonio Histórico Andaluz. El objetivo es crear un sistema de acceso y gestión uniforme a la información dispersa en diversas bases de datos que han sido creadas por grupos de trabajo independientes. Para ello, se propone un modelo de arquitectura, basado en XML, que permite el acceso uniforme a bases de datos heterogéneas y distribuidas.

1 Introducción

El Instituto Andaluz de Patrimonio Histórico (IAPH) es una institución dependiente de la Dirección General de Bienes Culturales de la Consejería de Cultura de la Junta de Andalucía. El IAPH se encuentra dividido en Centros de Investigación y Desarrollo en los que se cataloga, gestiona, difunden y restauran los bienes patrimoniales de Andalucía. Uno de sus centros es el Centro de Documentación en el que se desarrollan las bases de datos necesarias para la gestión y difusión del Patrimonio Histórico. En este trabajo se analizarán las necesidades que le han surgido a este Centro para desarrollar un Sistema de Información para integrar los datos de los bienes patrimoniales y la solución arquitectónica basada en bibliotecas digitales.

2 El Centro de Documentación

Dentro del Centro de Documentación existen fundamentalmente cuatro áreas temáticas de investigación y catalogación: Área de Arqueología, Área de Arquitectura, Área de Etnología y Área de Bienes Muebles. En cada una de estas áreas se han ido desarrollando aplicaciones que puedan soportar las necesidades de gestión de información. Estas cuatro áreas se subdividen a la vez en dos grandes grupos: Bienes Muebles y Bienes Inmuebles, tal y como se muestra en la figura 1. En esta figura se indican, además, otro grupo de bases de datos que podríamos llamar comunes por el hecho de que sus datos son compartidos por todas las bases de datos temáticas del Centro. Por un lado, encontramos la aplicación Tesoro de Patrimonio Histórico, que almacena los términos permitidos y aceptados por los investigadores y

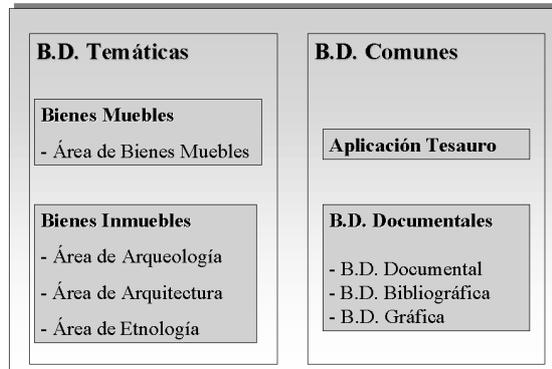


Figura 1: Estructura Informática del Centro de Documentación

catalogadores en una serie de campos como son la tipología o el periodo histórico del bien. Por otro lado, la información que se recoge sobre los bienes no es información puramente relacional: su estado de conservación, su denominación, sus autores, etc. Existe otro tipo de información que es la información documental. Las bases de datos documentales (figura1) se agrupa en tres: Gráfica, donde se almacenan los datos y las imágenes digitalizadas de los bienes; Bibliográfica, donde se almacenan los datos de los libros que hacen referencia o que estudian a un bien; y Documental, en la que se recogen documentos y artículos relacionados con el bien. Una imagen, un documento o un libro, puede hacer alusión a varios bienes a la vez, por lo que se recogen en bases de datos independientes que se relacionan con los bienes a través de los códigos tanto de la documentación como de los bienes.

3 Problemas a la estructura

A la hora de realizar las catalogaciones, esta estructura resulta muy conveniente. Un arqueólogo trabajará exclusivamente con la base de arqueología y los etnólogos con la base de datos de etnología. Sin embargo, no hay que olvidar que el fin último del Centro de Documentación es la de catalogar, documentar y difundir de forma ágil la información de los bienes. Es decir, facilitar consultas variadas de forma eficiente. A la hora de la catalogación, hay bienes que tienen tanto interés para los arqueólogos, como para los arquitectos y para los etnólogos. Así un bien puede estar en las tres bases de datos temáticas. Incluso podrían haberse denominado de diferente forma en dichas bases de datos y tener diferente código. Téngase en cuenta que las catalogaciones son independientes unas de otras. Además, en cada base de datos temática, hay información específica de la rama. Pero, por otro lado, también hay información que es común a todos, como la tipología o el período histórico. En resumen, hay parte de la información que se encuentra duplicada en las distintas bases de datos temáticas, pudiéndose así provocar los clásicos problemas de consistencia.

Todos estos problemas, encaminan el estudio hacia el desarrollo de una aplicación que integre la información de todas estas bases de datos en una sola [2]. Esta aplicación se denomina Sistema de Información Integrado de Patrimonio Histórico Andaluz.

4 Propuesta de arquitectura

La arquitectura propuesta se trata básicamente de un modelo para arquitecturas cliente-servidor [1]. Consta de cuatro capas, correspondiendo cada una a un proceso de transformación del flujo de información: *interfaz de usuario*, *sistema interrogador*, *sistema unificador* y *almacenamiento de datos*. La 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. El sistema interrogador traslada las peticiones del usuario a un lenguaje de consulta de datos. El sistema unificador desglosa las peticiones del usuario para recoger la información de las bases de datos correspondientes. Por último, el almacenamiento de datos almacena la información de forma distribuida en múltiples y variadas bases de datos.

De esta forma el usuario tiene una visión global e integrada de toda la información disponible, sin que eso suponga reestructurar todas las bases de datos sectoriales. Nosotros proponemos el uso de XML (*eXtensible Markup Language*) [3]. XML es un lenguaje de marcado, es decir, es un lenguaje que utiliza marcas para delimitar el comienzo y el final de un determinado tipo de dato. Este lenguaje 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 va a contener un determinado archivo XML, definimos el DTD (*Document Type Definition*). Un archivo DTD es un documento que complementa a un archivo XML describiendo los elementos que pueden aparecer en el archivo XML y la organización estructural que tienen. Junto a la definiciones de tipos de datos descritos en un documento DTD y a los datos contenidos en un documento XML, utilizamos documentos que reciben el nombre de *hojas de estilo* que nos permiten mostrar de forma determinada la información del documento XML. Para el caso de XML se han definido los formatos CSS (*Cascade Style Sheet*) y XSL (*eXtensible Stylesheet Language*). La propuesta de arquitectura se basa en el esquema de la figura 2.

Vemos que en el nivel del servidor aparecen dos bases de datos: las denominadas “instituto”, que son las bases de datos sectoriales descritas en la figura 1 y que mantendrían su estructura, y las denominadas “DTD-XSL-CCS”, que incluyen los metadatos y las hojas de estilo. El nivel intermedio lo constituye el mediador, que opera con el nivel inferior a través de un ORB (*Object Request Broker*) que independiza el acceso a las bases de datos del sistema de consulta. En este nivel también se encuentra el “XML middleware”. El XML middleware se encarga de recibir las peticiones realizadas desde el nivel cliente y de traducirla de manera que el nivel servidor las entienda. 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. 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).

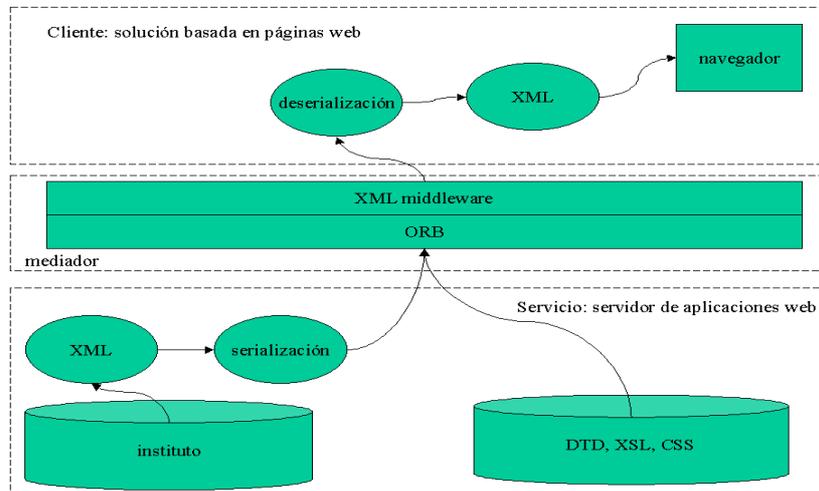


Figura 2: Arquitectura

Los procesos de serialización y deserialización que aparecen en los niveles servicio y cliente, tienen como fin facilitar el flujo de información entre capas. El usuario de la aplicación es totalmente ajeno a la complejidad interna del sistema. Así, un usuario 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. El mediador se encargará de hacer que el formato de estas consultas sea el adecuado e ir a la base de datos que pueda dar los resultados solicitados.

5 Conclusiones

En este documento se ha analizado un problema real, el del IAPH y una posible solución arquitectónica en la que el problema se resuelve tomándolo como un sistema para el tratamiento de bibliotecas digitales. Dicha propuesta ha sido tomada por el IAPH y está siendo actualmente aplicada. Sin embargo, no se está desarrollando el sistema en su totalidad desde el principio. El problema está siendo afrontado por fases, estando actualmente en una fase inicial. En esta fase sólo se incluyen una parte de los datos que debe contemplar la aplicación, pero poco a poco, se irán incluyendo todas esas necesidades.

6 Referencias

1. Brisaboa, Nieves. M.J. Durán, C. Lalín, J.R. López, J.R. Paramá, M.R. Penabad, A.S. Places, "Arquitectura para Bibliotecas Virtuales", 1999.
2. Witten, I. Craig Nevill-Manning, Sally Jo Cunningham. "Building a Digital Library for CSR. Technical Issues". Proc. of the 19th Australasian Computer Science Conf. 1996.
3. World Wide Web Consortium. www.w3c.org, 2000.