

APLICACIÓN DE LOS SISTEMAS DE TRATAMIENTO DE BIBLIOTECAS DIGITALES A LA GESTIÓN DEL PATRIMONIO HISTÓRICO

J. M. Cordero, M. J. Escalona, J. Torres, M. Mejías, R. M. Gasca*

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 hipertexto, 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 una solución a un problema real que se plantea en la gestión del 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.

Palabras clave: Bibliotecas digitales, patrimonio histórico, arquitectura software, XML.

I. 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, los Institutos de Patrimonio están desarrollando sistemas de información de patrimonio histórico (SIPH) con los que difundir la información que recopilan y con la que catalogan el patrimonio histórico. Sin embargo, en la actualidad, los SIPH no son sistemas sencillos puesto que mezclan una serie de requisitos que los transforman en sistemas complejos 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

* Departamento de Lenguajes y Sistemas Informáticos, Universidad de Sevilla. {jtorres@lsi.us.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.

II. PLANTEAMIENTO DEL PROBLEMA

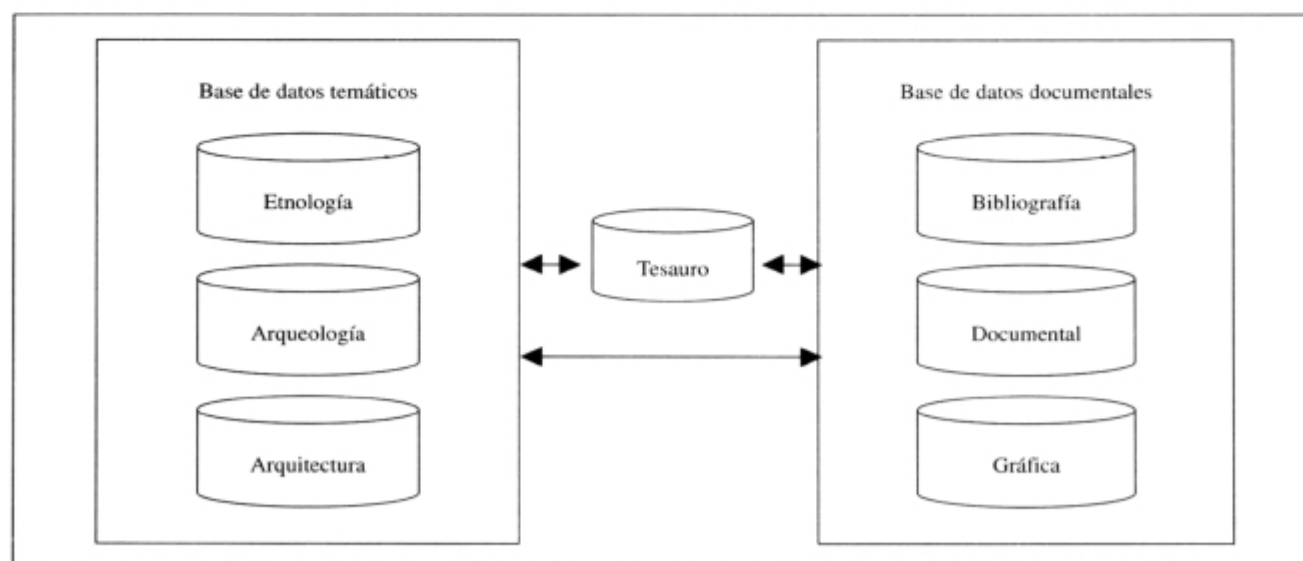
En este apartado vamos a describir el problema que vamos a abordar a lo largo del documento, detallando sus módulos y problemas fundamentales.

Un sistema de información de patrimonio histórico inmueble (SIPHI) guarda la información referente a los bienes inmuebles, por ejemplo, de una comunidad autó-

noma. 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 los Institutos de Patrimonio suele existir un sistema de información sectorial por cada una de las ramas: sistema de información de patrimonio arquitectónico, sistema de información de patrimonio arqueológico y sistema de información de patrimonio etnológico. La información que en ellos se almacena podemos agruparla por módulos: módulo de identificación y localización, módulo de descripción, módulo de análisis, módulo de conservación y el módulo de protección. Parte de los datos que se incluyen en estos módulos están normaliza-

Figura 1
BASES DE DATOS COMPONENTES DEL SIPHI



dos, pudiendo tener solo los valores que se encuentren en un almacén de términos que denominaremos tesoro.

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 sistemas diferentes para cada área que dé soporte a cada una de ellas, cuando hay bienes que son catalogados por más de un área temática, 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 tesoro de términos que lleva asociado. En la gestión de patrimonio es común el hecho de que exista una aplicación tesoro que normaliza los términos que pueden usar los catalogadores. Así, existen una serie de campos, como podrían ser la tipología, el estilo o las actividades para las que se destina (o se destinaba) el bien, que solo pueden tomar valores concretos del tesoro.

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, etcétera, 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 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.

III. JUSTIFICACIÓN DEL USO DE BIBLIOTECAS DIGITALES

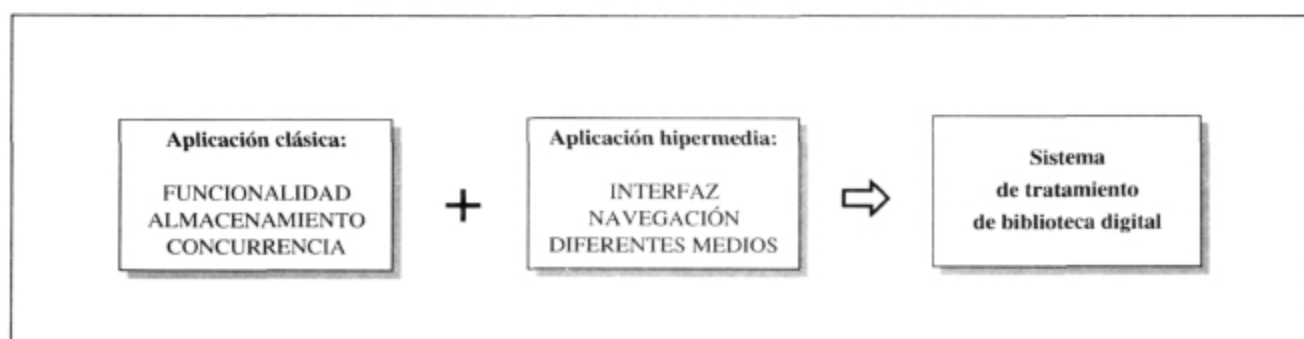
Un SIPH, como podemos concluir del apartado anterior, tiene una serie de característica que lo hace diferente de un sistema de información tradicional. En él se mezclan las necesidades de integración de elementos multimedia, con la posibilidad de navegar por la aplicación y con importantes

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

Figura 2

RELACION ENTRE LOS SISTEMAS DE TRATAMIENTO DE BIBLIOTECAS DIGITALES, LAS APLICACIONES CLÁSICAS Y LAS APLICACIONES MULTIMEDIA



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 para la gestión del patrimonio histórico.

IV. DEFINICIÓN DE UNA ARQUITECTURA

No existen actualmente modelos de arquitectura para el diseño de sistemas de tratamiento de bibliotecas digitales (Lesk, 1997), en parte por la novedad de las mismas y en parte por la falta de unanimidad en la definición de los medios que pueden integrarlos. La tendencia actual es tomar como referencias las arquitecturas ya existentes para el desarrollo de bases de datos, modificándolas con los modelos utilizados en el diseño de sistemas multimedia y sistemas interactivos. Pero la influencia cada vez mayor de Internet, tanto en el desarrollo de nuevos medios digitales como en la forma y capacidad de divulgación de los mismos, está impulsando la descripción de nuevos

modelos de arquitectura que reflejen las peculiaridades de Internet como sistema distribuido abierto. En este sentido los soportes de las bibliotecas se caracterizan por estar dispersos en la red, por estar implantados en sistemas heterogéneos y por la ausencia de estándares para el acceso a los mismos. Similares características presentan los operadores de las bibliotecas digitales.

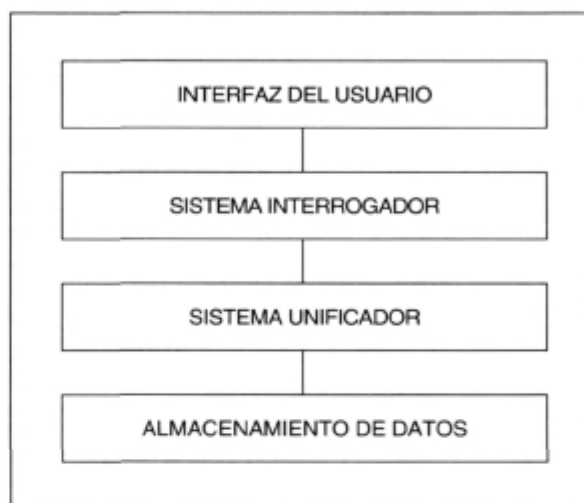
Nuestra propuesta de arquitectura contempla los siguientes requisitos:

- Los medios digitales disponen de algún mecanismo de descripción de la herramienta de observación o percepción de los mismos.
- Los soportes de los medios digitales son transparentes para el usuario.
- El acceso a la información se realiza de forma distribuida.
- No están predefinidos los criterios de búsqueda de información.

Por tanto, el modelo que proponemos contempla los siguientes aspectos:

- 1) Captura de información.
- 2) Almacenamiento.
- 3) Clasificación.
- 4) Interfaces para la presentación de resultados.

Figura 3
ARQUITECTURA GENÉRICA



- 5) Distribución del contenido de la biblioteca al usuario final.
- 6) Administración y control de acceso.

Una de las consideraciones que hemos tenido en cuenta en el modelo es la utilización de software y tecnologías ya existentes. Con esto se pretende que la configuración de un sistema construido siguiendo nuestro modelo sea lo más abierta posible. El modelo es también escalable, es decir, permite la incorporación de nuevas bases de datos y de nuevos medios digitales.

En las arquitecturas existentes (Witten, 1996) se destacan cuatro capas diferenciadas (figura 3):

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.

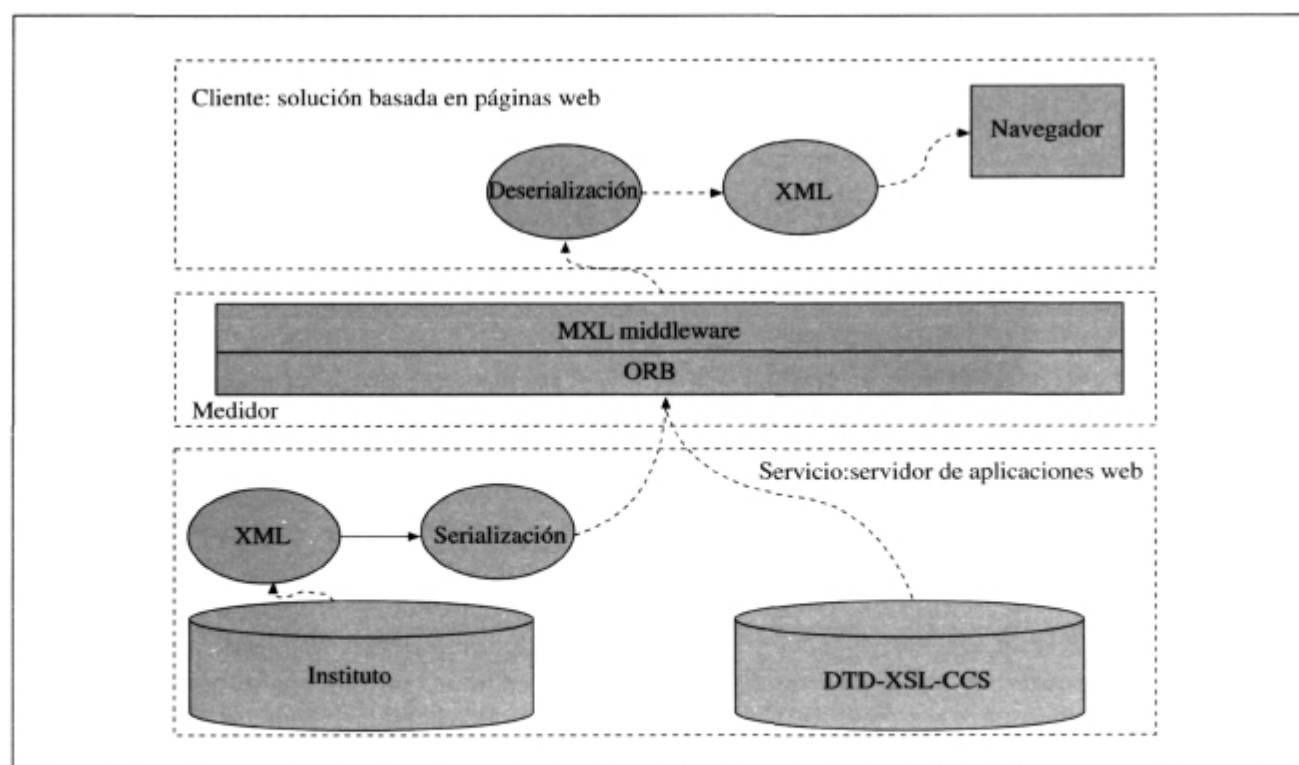
2. Sistema Interrogador: Traslada las peticiones del usuario a un lenguaje de consulta de datos.
3. Sistema Unificador: Desglosa las peticiones del usuario para recoger la información de las bases de datos correspondientes.
4. Almacenamiento de Datos: Almacena la información de forma distribuida en múltiples y variadas bases de datos.

Aceptando esta división genérica como idónea, es necesario encontrar un mecanismo que permita la distribución de la infor-

mación entre los niveles y que además esté normalizado. Este mecanismo es la utilización de XML (eXtensible Markup Language), propuesto por World Wide Web Consortium (WWW, 2000). XML 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.

Como complemento a los archivos XML generados utilizamos un documento DTD

Figura 4
MODELO DE LA ARQUITECTURA
PROPUESTA



(Document Type Definition) con la descripción de los elementos que pueden aparecer en un archivo XML y su organización.

Para las presentaciones de las consultas, 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). Para el uso interactivo del sistema tratamos con las entidades implicadas en el proceso de presentación y con los diálogos entre usuario y sistema, por lo que el uso de los modelos de objetos como DOM (Document Object Model) nos permiten, desde lenguajes como JavaScript, la manipulación de los objetos presentes en el documento visualizado por el navegador.

Todo este planteamiento se concreta en la figura 4, que recoge un enfoque cliente-servidor, donde se distinguen tres niveles diferenciados. 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. 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).

El nivel intermedio 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, que aglutina a los sistemas in-

terrogador y unificador de la arquitectura genérica.

Por último, en el nivel servicio se encuentran dos tipos de bases de datos, las denominadas "instituto", que hacen referencia a aquellas que contienen la información objeto de tratamiento y propiedad del Instituto de Patrimonio, 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.

La integración de los dos tipos de información almacenada en las diferentes bases de datos presentes en el nivel servicio la realiza el XML middleware existente en el mediador.

Por otro lado, 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.

V. CONCLUSIONES

En este trabajo se ha planteado la problemática que rodea a la definición de sistemas de gestión 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.

Siendo conscientes de que este problema se plantea en todos los Institutos de Patrimonio nacionales y en todos los centros de gestión de patrimonio, hemos centralizado la solución basándonos en el Instituto Andaluz de Patrimonio Histórico. Analizada la propuesta realizada en este entorno, se puede concluir su aplicabilidad y la validez de los resultados obtenidos.

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.

BIBLIOGRAFÍA

- BRISBOA, N.; DURÁN, M. J.; LALÍN, C.; LÓPEZ, J. R.; PARAMÁ, J. R.; PENABAD, M. R., y PLACES, A. S. (Brisaboa, 1999): "Arquitectura para Bibliotecas Virtuales, 1999.
- LESK, M. (Lesk, 1997): *Practical Libraries: Books, Bytes & Bucks*. Morgan Kaufmann. 1997.
- MARTÍNEZ, J. M.; HILERA, J. R.; MARTÍNEZ, J. J., y GUTIÉRREZ, A. (Martínez, 1997): "Orientación a Objetos en la Documentación Hipermedia", Universidad de Alcalá de Henares, Departamento de Ciencias de la Computación.
- WITTEN, I.; NEVILL-MANNING, C., y JO CUNNINGHAM, S. (Witten, 1996): *Building a Digital Library for Computer Science Research: Technical Issues*. Proceedings of the 19th Australasian Computer Science Conference. 1996.
- WORLD WIDE WEB CONSORTIUM. (WWW, 2000) www.w3c.org, 2000.