

Generación de documentación hipermedia en Internet a partir de información multimedia en Bases de Datos

José Antonio Gutiérrez De Mesa
José Ramón Hilera González

RESUMEN

Se describe un método para el acceso a información multimedia almacenada en bases de datos situadas en servidores Web, presentándola al usuario en forma de páginas HTML que se crean dinámicamente cuando son demandadas desde cualquier navegador. Se evitan así los problemas de mantenimiento y seguridad que supone la gestión de un gran número de páginas Web en los servidores, aprovechando las facilidades ofrecidas al respecto por los Sistemas de Gestión de Bases de Datos Relacionales existentes.

INTRODUCCIÓN

El término HIPERMEDIA, combinación de los conceptos HIPERtexto y multiMEDIA, hace referencia a una tecnología de construcción de (hiper)documentos que permite a la audiencia de los mismos encontrar fácilmente la información que realmente necesitan, a través de hiper-enlaces establecidos entre los diferentes elementos de información multimedia (texto, sonido, imagen fija, imagen en movimiento) que conforman los documentos (1). La aparición del entorno WEB en el ámbito de Internet ha permitido que se pueda publicar y acceder a documentación hipermedia de forma extraordinariamente sencilla y con un coste muy reducido, mediante la utilización de servidores Web para almacenar los documentos publicados, y navegadores (como *Microsoft Explorer* o *Netscape Navigator*) para acceder a su contenido multimedia: mediante la visualización de texto e imágenes, la audición de sonidos y la reproducción de vídeo; y para facilitar la navegación entre documentos a través de hiper-enlaces.

Los documentos puestos a disposición de los usuarios de la red normalmente no forman parte de bases de datos en el sentido estricto del término, sino que se suelen distribuir en forma de páginas Web registradas en archivos independientes almacenados en servidores de información, conteniendo estas páginas los enlaces que, con la ayuda de un navegador, permitan que un usuario "salte" de un documento a otro, aun cuando estos documentos estén situados en diferentes computadores comunicados a través de la red.

Cuando una persona u organización desea ofrecer un volumen importante de documentación a los usuarios a través de Internet, se plantean algunos problemas cuya resolución dependerá de si se va a adoptar un enfoque tradicional, basado en un sistema de ficheros (páginas Web), o un enfoque fundamentado en la utilización de un sistema de gestión de base de datos. Estos problemas son los siguientes:

- La gestión del almacenamiento de los documentos
- La implementación de los hiper-enlaces entre documentos
- El mantenimiento de la documentación
- El mantenimiento de los hiper-enlaces
- La seguridad de la información

El primero de estos problemas se refiere a la forma de organizar la información en los soportes físicos de acceso directo (normalmente, discos magnéticos) para almacenar de forma óptima los documentos. Hacerlo mediante ficheros sin más estructura que la permitida por el Sistema Operativo, basada en directorios o carpetas, cuando su volumen es muy grande no parece ser lo más adecuado, si se tiene en cuenta la existencia de Sistemas de Gestión de Bases de Datos, especialmente creados para facilitar el registro y la recuperación de grandes cantidades de información con un coste de mantenimiento reducido.

Frente al segundo problema, se puede adoptar simplemente la implementación de los enlaces en el interior de los documentos o complementarlo con las posibilidades que ofrece la gestión de índices que incorpora cualquier Gestor de Bases de Datos.

El mantenimiento de la información que se va a poner a disposición de los usuarios de la red es un importante problema a considerar, ya que actualizar grandes volúmenes de información supondrá un esfuerzo considerable si ésta se almacena directamente a nivel de presentación, es decir, en forma de páginas Web incluidas en archivos y codificadas mediante un lenguaje de marcado como HTML. Es más eficaz adoptar una estructura de la información en forma de base de datos, relacional u orientada a objetos, que, en principio, no considere la forma en que el usuario visualizará tal información (como páginas Web).

En cuanto al mantenimiento de los hiper-enlaces, si se utilizan las facilidades de gestión de índices de un Sistema de Bases de Datos, el propio mantenimiento de la información implicará la actualización automática de los índices de acceso a documentos relacionados.

Frente al problema de la seguridad, es necesario considerar dos aspectos: la seguridad en la información documental almacenada y la seguridad en la información que se transmite por la red. Si se utiliza un Sistema de Gestión de Bases de Datos, la seguridad del almacenamiento, en cuanto al proceso de transacciones y a la integridad referencial, la garantiza el propio gestor de la base de datos. La seguridad en la red se puede solucionar con el empleo adicional de unas claves públicas entre el software de acceso a la base de datos en el servidor Web y el navegador.

GENERACIÓN DE PÁGINAS WEB VIRTUALES

Como se ha indicado en el apartado anterior, el almacenamiento de la información que contienen los documentos en una Base de Datos puede ofrecer ventajas frente a su almacenamiento en forma de ficheros que contienen cada uno de ellos una página Web. Como Sistema de Base de Datos se puede adoptar uno de tipo relacional, el más ampliamente utilizado en la actualidad, aunque el futuro apunta hacia los denominados Sistemas Orientados a Objetos como los más eficientes para gestionar datos de muy diferente naturaleza, como es el caso de los datos multimedia (2). No obstante, muchos sistemas relacionales permiten utilizar campos BLOBs (*Binary Large Objects*) en las Bases de Datos Relacionales para registrar información de tipo multimedia, como sonidos o imágenes.

Para la generación de páginas Web de forma dinámica a partir de la información almacenada en una base de datos, además del programa gestor de dicha base de datos, normalmente comercial, es necesario disponer de otro programa que atienda las peticiones que realizan los usuarios vía Internet y construya, en tiempo real, la página Web "virtual" correspondiente, tomando de la base de datos la información necesaria. Para independizar este segundo programa del gestor comercial de la Base de Datos utilizado, se puede añadir un componente

software intermedio, o interfaz, que normalice las relaciones entre ambos programas, de tal forma que si se cambia de gestor de base de datos (por ejemplo, Oracle por Informix, Informix por Microsoft SQL Server, etc.), no sea necesario introducir cambios en el programa que accede a la base de datos para construir páginas Web. En el ámbito de los Computadores Personales (PCs) la interfaz de este tipo más utilizada es la conocida como ODBC (*Open Data Base Connectivity*).

La estructura de los componentes necesarios para la generación de páginas Web virtuales a partir de la información multimedia contenida en Bases de Datos es la que se muestra en la figura 1.

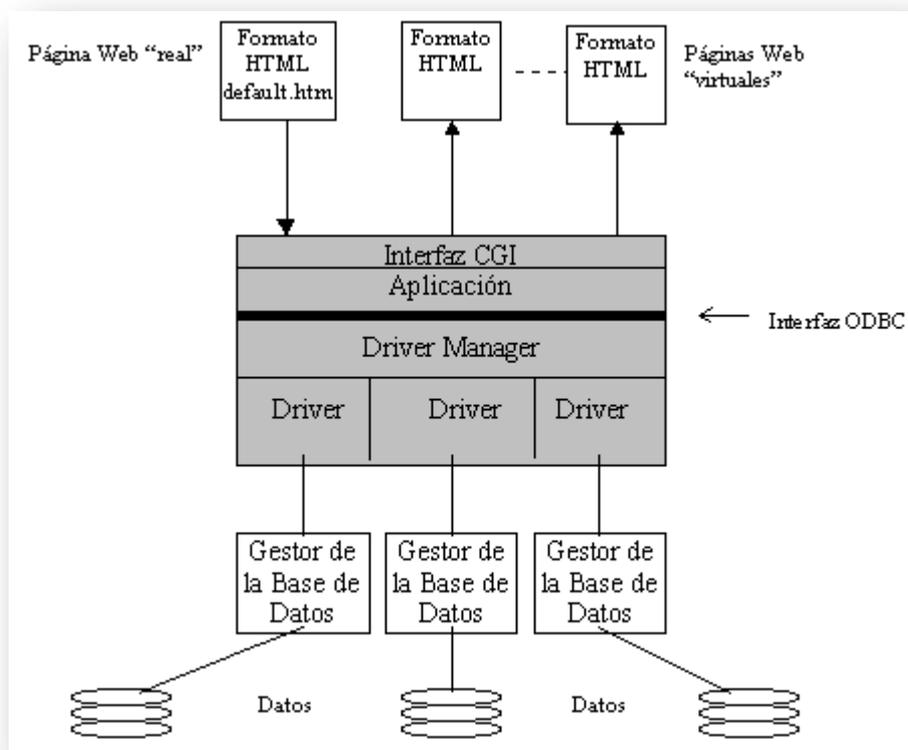


Figura 1. Componentes para la generación de páginas Web "virtuales"

Como puede observarse en la ilustración, debe existir una página Web "real", almacenada en el servidor Web en forma de archivo ".htm" con formato HTML. Esta página normalmente será la denominada "default.htm", que es la que se "abre" por defecto cuando un usuario accede mediante un navegador a una dirección de "sitio" Web en Internet sin especificar un nombre de página: por ejemplo, es equivalente acceder a la dirección www.cc.alcala.es que a www.cc.alcala.es/default.htm.

La página Web "real" es de presentación, ofreciendo al usuario el acceso a otras páginas que no existen como tales en el servidor, sino que serán creadas en tiempo real con la información

de la Base de Datos. Para ello, los enlaces en la página de presentación no hacen referencia a ficheros HTML, sino a un programa que al ejecutarse origina la ejecución de la aplicación que "construye" dinámicamente la página "virtual" correspondiente. De esta forma, en lugar de incluir un hiper-enlace del tipo:

```
<A HREF="Nombre fichero que contiene página Web">
```

(Por ejemplo:)

se utilizaría otro que originase la ejecución del programa de creación de la página:

```
<A HREF="Nombre del programa">
```

(Por ejemplo:)

En la figura 1 también aparece un componente denominado "Interfaz CGI" (*Common Gateway Interface*); se trata de un programa que permite relacionar aplicaciones desarrolladas en un determinado lenguaje de programación (por ejemplo, C++, Java, etc.) con el mundo Web. Además de esta interfaz que comunica la aplicación encargada de crear las páginas virtuales con Internet, también debe existir otra interfaz que permita su relación con el mundo de las Bases de Datos. Esta interfaz puede ser, como se ha mencionado anteriormente, ODBC, que incluye, como puede apreciarse en la figura, un *Drive Manager*, programa encargado seleccionar el *driver* adecuado en función del Sistema Operativo (Windows, Unix, etc.) y del Sistema de Gestión de Base de Datos (Oracle, Informix, etc.) utilizados. Un *driver*, en este contexto, es el programa que procesa todas las órdenes de acceso a la base de datos, normalmente expresadas en un lenguaje de consulta denominado SQL (*Structured Query Language*), y las convierte en comandos que pueda "entender" el Gestor de la Base de Datos utilizada.

APLICACIÓN PRÁCTICA

En el Departamento de Ciencias de la Computación de la Universidad de Alcalá se ha llevado a cabo un proyecto piloto para comprobar este método de generación de documentación hipermedia en Internet. Para ello, se ha creado una Base de Datos Relacional, utilizando la herramienta comercial *Microsoft SQL Server*, con información multimedia sobre los más famosos pintores españoles. En esta base, además de textos, se han incluido fotografías de calidad de sus principales cuadros, explicaciones audibles en forma de sonido digitalizado sobre autores y cuadros, música de la época, y vídeo digital.

Se han creado únicamente dos páginas Web "reales" en formato HTML, ambas de presentación, con el mismo contenido, aunque una de ellas en español y la otra en inglés. En la página de presentación existen hiper-enlaces, a través de una interfaz CGI, con la aplicación que crea las páginas Web "virtuales" accediendo a la base de datos mediante órdenes SQL, siendo esto transparente al usuario, y mostrándole su contenido (imágenes, texto, etc.) como si de páginas Web se tratara; páginas que se destruyen cada vez que se solicita una nueva consulta, aunque mediante su navegador, el usuario puede examinar su código HTML y almacenarlas en su computador para su propio uso.

CONCLUSIONES

El método de generación de páginas Web "virtuales" descrito ofrece ventajas evidentes que se han puesto de manifiesto en el proyecto piloto sobre pintores españoles que se ha desarrollado. Como ventaja principal podemos destacar la facilidad del mantenimiento del gran volumen de información utilizada, al estructurarse en forma de base de datos, además de poder aprovechar todas las facilidades ofrecidas por los gestores de bases de datos comerciales en lo que se refiere a la seguridad de la información y a la integridad de los índices utilizados para implementar los hiper-enlaces de navegación.

En el futuro trataremos de mejorar la eficacia de esta forma de gestionar documentación hipermedia, utilizando bases de datos orientadas a objetos en lugar de relacionales (3), aplicando técnicas de modelado conceptual para el diseño de la estructura de la información multimedia (objetivo del presente trabajo), y explotando la posibilidad de diálogo entre servidores para habilitar hiper-enlaces entre distintas bases de datos distribuidas en la red.

BIBLIOGRAFÍA

MARTÍNEZ, José Manuel; HILERA, José Ramón; MARTÍNEZ, Javier y GUTIÉRREZ, José Antonio. (1996): Orientación a Objetos en la Documentación Hipermedia. II Jornadas sobre Tecnologías de Objetos. Madrid, Asociación de Técnicos en Informática, SIMO TCI, 1996, 49-54.

SIMON, A.R. (1995): Strategic Database Technology: Management for the Year 2000. Morgan Kaufmann Publishers, 1995.

GUTIÉRREZ, José antonio; HILERA, José Ramón y MARTÍNEZ, Javier (1996): Requisitos Funcionales de una Base de Datos Multimedia para la Optimización de los Planes de Consulta. I Jornadas sobre Inteligencia Artificial, Control y Sistemas Expertos, Alcalá de Henares, Julio, 1996, 139-144, Universidad de Alcalá, 1996.

MARTÍNEZ, José Manuel y HILERA, José Ramón (1998): "Modelado de documentación multimedia e hipermedia". Cuadernos de Documentación Multimedia, Madrid, Nº 6-7. <<http://www.ucm.es/info/multidoc/multidoc/revista>>.

NOTAS

(1) Martínez, J. M; Hilera, J. R.; Martínez, J. y Gutiérrez, J.A.; 1996.

(2) Simon, 1992

(3) Gutiérrez, Hilera, Martínez, 1996.