

PROYECTO DIOSCÓRIDES

AURORA MIGUEL ALONSO
Directora del Proyecto

PILAR MORENO GARCÍA
Coordinación Técnica

J. CARLOS HERRÁIZ REGIDOR
Servicios Informáticos

I. INTRODUCCIÓN

La Universidad Complutense de Madrid y la Fundación Ciencias de la Salud han firmado recientemente un Convenio de colaboración para la puesta en marcha de un proyecto conjunto de investigación denominado Proyecto Dioscórides.

Este proyecto consiste en la creación de una «biblioteca electrónica» de fondo histórico biosanitario mediante la digitalización del material bibliográfico de los siglos XV a XVIII de las bibliotecas de las Facultades de Farmacia y Medicina, al que se incorporarán los ejemplares de esta temática existentes en otras colecciones bibliográficas de la Universidad.

El Proyecto Dioscórides responde al esfuerzo y colaboración, tanto de la Fundación Ciencias de la Salud, que proporciona equipamiento técnico y soporte informático, como de la Universidad Complutense de Madrid, que pone a disposición su riquísimo fondo bibliográfico y la infraestructura de personal, especialistas en fondo antiguo, informáticos y becarios, imprescindibles para sacar adelante un proyecto tan ambicioso.

La puesta en marcha de este proyecto supone evidentes ventajas, que se pueden resumir en:

- Una mejor conservación de las obras digitalizadas, no sólo de los ejemplares reproducidos, sino también, en muchos casos, del resto de las ediciones existentes (siempre en el ámbito histórico) en la Biblioteca de la Universidad.
- Una más rápida localización de los ejemplares, con visualización directa en pantalla y posibilidad de consulta remota de la colección y de reproducción en papel con una gran calidad.

- Un conocimiento más exhaustivo por parte del mundo científico y bibliotecario, cualquiera que sea su lugar de trabajo, de la riquísima colección de fondo histórico de temática biosanitaria existente en la Biblioteca de la Universidad Complutense de Madrid.

II. CARACTERÍSTICAS GENERALES DE LA DIGITALIZACIÓN.

CARACTERÍSTICAS DE NUESTRO PROYECTO

La digitalización electrónica consiste, básicamente, en capturar las imágenes de los documentos mediante un proceso de escaneo. Las imágenes así captadas se almacenan en un soporte óptico o magnético. Más tarde, esta imagen se transforma para su compresión e interpretación.

Los parámetros fundamentales en el proceso de digitalización son la resolución (número de puntos explorados por unidad de longitud) y el número de niveles de gris en que se cuantifica la intensidad de luz reflejada por el documento. Ambos parámetros condicionan, por una parte, la obtención de imágenes digitales de suficiente calidad y, por otra, el almacenamiento total requerido.

Por las características del fondo a digitalizar y por los objetivos del sistema establecidos (la mayor parte del fondo con el que se va a trabajar se compone de material impreso en blanco y negro) se han elegido unos parámetros de digitalización de 400 dpi. de resolución. Esto permite asegurar una buena calidad de imagen, legibilidad de los caracteres y, al mismo tiempo, no excederse en las necesidades de espacio digital.

Las imágenes digitalizadas utilizan los algoritmos de compresión del grupo IV del Comité Consultivo Internacional Telegráfico y Telefónico (CCITT) de la Organización de las Naciones Unidas, desarrollados para reducir la duración de la transmisión de imágenes a través de los circuitos telefónicos. Su codificación se realiza de acuerdo con el formato TIFF (Tagged Image File Format) encargado de proporcionar el nivel más alto de normalización necesario para permitir el intercambio de imágenes.

III. CARACTERÍSTICAS DE LA APLICACIÓN

El desarrollo de la aplicación ha sido encargado por la Fundación Ciencias de la Salud a la empresa KEON, que trabaja en un diseño *ad hoc* para este proyecto. En el momento actual un analista de KEON trabaja a tiempo completo en la puesta a punto del sistema, y en la ayuda y resolución de los problemas que van surgiendo.

Se ha diseñado una aplicación basada en sistemas abiertos con una arquitectura *cliente-servidor*. Los sistemas abiertos son aquellos sistemas informáticos

que, además de incluir *software* y *hardware*, están unidos en redes de comunicaciones. En estos sistemas se ejecutan aplicaciones que ofrecen información a usuarios ubicados en distintos puntos de la red a través de modelos de presentación de datos o *interfaces* de usuarios.

La arquitectura *cliente-servidor* se entiende como un modo de cooperación entre dos sistemas informáticos. Uno de ellos, el servidor (un ordenador) realiza el trabajo y envía los resultados al cliente o clientes (otro ordenador). La unión entre ambos se realiza a través de una red de comunicaciones.

A su vez, el sistema desarrollado permite la conexión entre la base de datos bibliográfica de la Biblioteca de la Universidad Complutense, gestionada por LIBERTAS y el archivo de imágenes digitalizadas. Esta fórmula permitirá al investigador pasar de un modo natural de la búsqueda bibliográfica a la visualización de libro digitalizado que desea consultar. Para esta consulta, el usuario del sistema dispondrá de una ayuda *online* sobre el uso del mismo.

IV. ORGANIZACIÓN Y ACCESO DE LOS LIBROS EN EL ARCHIVO ÓPTICO

El sistema permite organizar el libro en diferentes módulos o partes sustanciales del mismo. El investigador podrá acceder tanto a las imágenes secuenciales que forman el libro como, directamente, a las partes diferenciadas que lo integran:

- Lomo.
- Cubiertas.
- Guardas.
- Portadas.
- Preliminares.
- Capítulos o Libros.
- Láminas.
- Índices Generales.
- Índice de Láminas.
- Páginas numeradas¹.

Un apartado especial en la recuperación de la información es la posibilidad de acceder directamente a las láminas de una obra, o recuperando por artista, materia y año/período, al conjunto de láminas que existan en los libros escaneados (lámina 1).

¹ El sistema respeta la paginación real del documento, que en el caso del fondo antiguo es variada y de gran complejidad.

Ejemplo 1 (Fig. 1)

Libro:

Lomo
 Cubierta
 Guarda
 Páginas:
 Portada
 Preliminares
 Capítulo:
 Índice General
 Índice I. Ánimas
 Colofón
 I. Ánima:

Lámina
Autor:
Materia:
Período: -

Diálogo de la factura en
latín y en francés.

<p><i>En ces livres bergères fructes Lectres parcellées et fructes Doctores chose haulte poudre bergeres, quant l'on garder Dre d'empes, les moustres ou haultes Lectres bergeres font en poise</i></p>	<p><i>Donc puis en se peut poudres. Just au poise, et par le fait Et quel est le poise d'empes Et en se bergeres commencent que de se peut moule moule Lectres parcellées et fructes Et apres la table d'empes On les fait le fait parcellées</i></p>
---	---

Ejemplo 2 (Fig. 2)

Libro:

Lomo
 Cubierta
 Guarda
 Página:
 Portada
 Preliminares
 Capítulo:
 Índice General
 Índice I. Ánimas
 Colofón
 I. Ánima:

Lámina
Autor:
Materia:
Período: -

Diálogo de la factura en
latín y en francés.

<p><i>En ces livres bergères fructes Lectres parcellées et fructes Doctores chose haulte poudre bergeres, quant l'on garder Dre d'empes, les moustres ou haultes Lectres bergeres font en poise</i></p>	<p><i>Donc puis en se peut poudres. Just au poise, et par le fait Et quel est le poise d'empes Et en se bergeres commencent que de se peut moule moule Lectres parcellées et fructes Et apres la table d'empes On les fait le fait parcellées</i></p>
---	---

Las múltiples posibilidades de recuperación buscan dar una atención personalizada. En concreto, hemos tratado de dar solución a las necesidades de tres tipos de usuarios:

a) Investigadores que quieren consultar el texto de la obra digitalizada. Este tipo de usuarios necesita ante todo recuperar visualmente el libro, página a página, un capítulo determinado... La aplicación permite la consulta directa de la portada, índices, del texto, dividido o no en capítulos o libros, etc.

b) Bibliógrafos, historiadores del libro, historiadores de colecciones y bibliotecas. Desean comprobar lo más directamente posible datos bibliográficos de la obra digitalizada, que se pueden recuperar en puntos muy concretos. El sistema les permitirá recuperar directamente la portada, el colofón, preliminares y guardas.

c) Historiadores del arte, iconógrafos, historiadores de la ciencia a través de la imagen, etc. Desean recuperar ante todo las series de imágenes de las obras, prescindiendo muchas veces, al menos en una primera etapa de la consulta, del texto de los datos bibliográficos más completos. Para este tipo de usuario se ha buscado una fórmula de recuperación que les permitirá resolver estas necesidades, con posibilidades de recuperación directa de las imágenes de un libro de un modo secuencial o por el nombre de un artista, de una materia o de un año o periodo artístico.

V. ORGANIZACIÓN

El equipo que lleva a cabo el proyecto está compuesto por personal multidisciplinario que responde a las distintas necesidades requeridas por un proyecto de estas características: nuevas tecnologías aplicadas a la documentación, gestión de fondo antiguo, soporte informático, etc.

En estos momentos se cuenta con el siguiente personal de la UCM:

Dirección del Proyecto	<i>Aurora Miguel Alonso</i>
Coordinación Técnica	<i>Pilar Moreno García</i>
Servicios Informáticos	<i>J. Carlos Herráiz Regidor</i>
Bibliotecarios de Fondo Antiguo	<i>Mercedes Cabello Martín</i> <i>Alberto Morcillo Ortega</i>
Becarios	<i>M.^a José Vilabella Tellado</i> <i>Patricia Esteban Gracia</i> <i>M.^a Teresa Cordero Muñoz</i>

Asimismo, el proyecto utiliza la configuración *hardware* & *software* que a continuación se detalla:

Un servidor compuesto por un Ordenador HP E25, 64Mb RAM y un Juke-Box HP 120T. El *software* utilizado por el servidor es una Base de Datos de tipo relacional Informix On Line.

La digitalización de imágenes se realiza mediante un escáner Kodak Imagelink 200 de cámara digital especialmente diseñado para fondo antiguo, de reciente aparición en el mercado, y que permite la recuperación de la imagen sin dañar la estructura del libro.

Las tareas de control, indexación y consulta se llevan a cabo con ordenadores personales HP-Vectra 486 de 16Mb utilizando un *software* Windows 3.11 TCP/IP LanWorkplace para DOS 4.2 de Novell.

Madrid, 20 de junio de 1995

