

## Creación de un sistema básico de escaneado de sumarios de publicaciones periódicas

Víctor M. Macías Alemán

Universidad de Las Palmas de Gran Canaria  
Biblioteca Universitaria

BIBLID 1133-9756 (2000)12 p. 59-63

El presente trabajo explica nuestra experiencia en la resolución de un problema común en las bibliotecas universitarias: cómo dar a conocer de forma eficaz al profesorado e investigadores el contenido de las publicaciones seriadas que se reciben.

Tradicionalmente, el sistema ha consistido en la fotocopia sistemática y su agrupamiento en cuadernillos de periodicidad variable, frecuentemente mensuales. Esto suponía un gran gasto de papel, retraso en su distribución y una obsolescencia acusada. Además, era imposible hacérselo llegar a todos los usuarios interesados, profesores e investigadores de nuestra Institución, por el alto coste que ello suponía.

Por todo ello y teniendo en cuenta los medios informáticos con los que es posible contar en la actualidad, sopesamos las diversas posibilidades existentes a nuestro alcance. La alternativa, en primera instancia, podía dirigirse a contratar los servicios de una empresa especializada en la digitalización documental. Sin embargo, aunque ya somos clientes a este respecto con la digitalización sistemática de la

prensa corriente canaria que se nos viene haciendo desde 1996, valoramos la necesidad de introducirnos por nuestros propios medios en este campo, a fin de empezar a conocer y desarrollar sistemas propios con los que resolver nuestras necesidades más puntuales. Para ello, una tarea sencilla y de bajo volumen de trabajo como la digitalización sistemática de sumarios de las publicaciones periódicas suscritas, se nos presentaba como ideal para introducirnos en esta técnica que cada día es más necesaria en el ámbito de las bibliotecas científicas.

### 1. SELECCIÓN DEL SISTEMA INFORMÁTICO

El sistema informático que deberíamos escoger para sustituir el procedimiento manual de difusión de los sumarios debía atender a tres requisitos: sencillez de puesta en marcha y manejo; rapidez en los resultados, y escaso coste económico, tanto en los equipos como en el número del personal necesario para atenderlo.

Para atender a ello, debíamos seleccionar un equipo de digitalización compuesto de: ordenador personal, escáner y software de digitalización y distribución.

### 1.1. HARDWARE

El ordenador personal elegido incorpora las siguientes características: monitor color de 21 pulgadas, tarjeta de vídeo de 16 Mb AGP, procesador Pentium III a 800 Mhz, 128 Mb SDRAM, 2 discos duros de 6 y 20,4 Gb respectivamente y grabadora de CD-ROM.

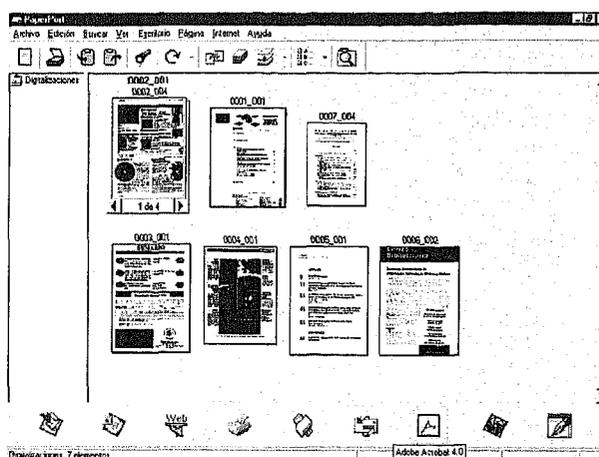
Con respecto al escáner, decidimos mantener el actual sistema de producción de los originales de sumarios, mediante el fotocopiado en papel formato A4 de todos ellos, a partir de los cuales trabajar. Por consiguiente, ello nos dirigía directamente a la selección de escáneres A4 de cama plana con alimentación automática.

Entre las marcas disponibles en el mercado seleccionamos una de las existentes con más garantías en el sector de este tipo de productos, Hewlett-Packard, por la diversidad de modelos adecuados para nuestro propósito, la existencia de accesorios de ampliación, las debidas garantías y la posibilidad de evolucionar, llegado el momento, hacia otros productos compatibles. El modelo elegido fue el de más alta capacidad de trabajo, el 6390C, de alta resolución (1200 dpi óptica, 1200 x 2400 dpi hardware), con conexiones SCSI y USB, alimentador automático de 25 hojas y que incorporaba además de serie otros accesorios, lo que podía permitir cara al futuro el escaneo de otro tipo de materiales en blanco y negro, en escala de grises o color, dentro de sus límites máximos de A4.

### 1.2. SOFTWARE

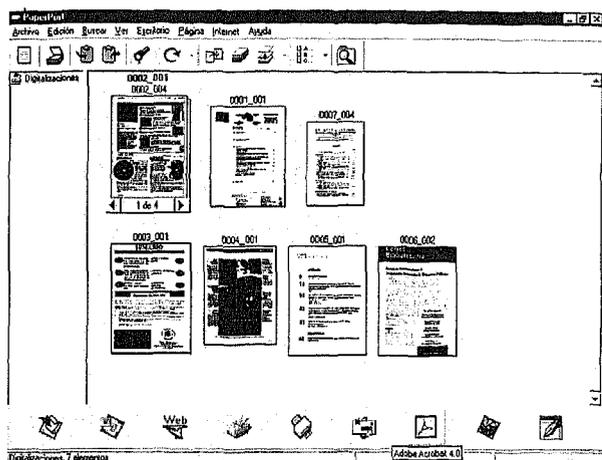
Con respecto al software existía la necesidad de contar con, al menos, dos programas específicos para los siguientes propósitos:

- Para la gestión del escaneado seleccionamos la aplicación PaperPort 6.5 Deluxe. Este programa permite corregir en el aspecto que deseemos la imagen escaneada. De la misma forma, permite apilar varias hojas escaneadas si todas ellas pertenecen al mismo sumario de un determinado ejemplar de publicación periódica. El resultado se puede automatizar y, tras ello, ser enviada la digitalización a la aplicación que se desee.



Sumarios escaneados con PaperPort 6.5 Deluxe

- Para la generación de los ficheros de las digitalizaciones de los sumarios que han de consultarse elegimos Adobe Acrobat 4.0. Ello fue debido a diversas razones, siendo la más destacada que el acceso a los sumarios digitalizados se realizaría a través de la Web, actualmente el estándar de facto de distribución de documentación digitalizada a texto completo a través de Internet lo constituye el formato propietario PDF (*Portable Document Format*), el cual es generado y gestionado con Adobe Acrobat. Por otro lado, este formato de ficheros es accesible independientemente de la plataforma que posea el usuario (Unix, PC, Macintosh...), del sistema operativo (Windows, OS/2, Linux...) y su interface está en la mayoría de los idiomas de ámbito internacional.



Sumarios convertidos a PDF con Adobe Acrobat 4.0

Entre otras utilidades, Acrobat incorpora además un OCR integrado que puede calificarse de excelente, por lo que en cualquier momento podemos convertir a texto editable las imágenes de los documentos. Además, el programa cliente para acceder a los documentos PDF es totalmente gratuito y disponible en multitud de lugares para su fácil descarga al ordenador del usuario.

A estos programas específicos unimos otros, como la confección de una base de datos en Microsoft Access para la gestión del inventario de las digitalizaciones realizadas, con la que poder conocer en todo momento el estado del trabajo realizado; o un cliente FTP para poder transferir los ficheros escaneados al servidor desde donde ofertar las imágenes de los sumarios a los usuarios interesados.

## 2. PROCEDIMIENTO DE TRABAJO

Para proceder debidamente elaboramos un manual de procedimiento que, aunque sujeto permanentemente a modificación –según que la experiencia y recomendaciones existentes nos fueran indicando en cada parte del trabajo por dónde continuar–, nos indique en todo momento las pautas a seguir en cada paso.

### 2.1. REALIZACIÓN DE LAS FOTOCOPIAS ORIGINALES

En primer lugar, fue enviada una circular a todas las bibliotecas sucursales generadoras de sumarios para que atendieran a una serie de requisitos de calidad en el fotocopiado de los originales que, posteriormente, recibiríamos nosotros para su digitalización. Estos criterios son:

- Realizar fotocopias originales, no fotocopias sacadas a partir de otras.
- Evitar cualquier tipo de márgenes oscuros.
- No realizar fotocopias a doble cara o dos en una misma cara, ya que no sería posible entonces el uso eficaz del alimentador de páginas.
- En caso de que no esté plenamente identificado en el sumario el título o fecha o volumen o número de la publicación, anotar estos datos con letra clara y mayúsculas en el reverso de la fotocopia a lápiz, a fin de identificarla y evitar que se transparente.
- Para aquellos casos excepcionales de publicaciones periódicas cuyo sumario esté ubicado en cubiertas o tipos de papel difíciles de fotocopiar con la debida calidad (papel satinado, colores oscuros de fondo), deberán teclear dicho sumario en procesador de texto, adjuntándonos la impresión al resto de las fotocopias normales. Deberá incluir el título de la revista, fecha, volumen y número.

### 2.2. CONTROL DEL TRABAJO REALIZADO

Obtuvimos en primera instancia, a través de la Hemeroteca Universitaria, el listado completo de las publicaciones seriadas suscritas, indicando Título, ISSN y la biblioteca sucursal en que se reciben.

A continuación creamos una base de datos en Microsoft Access con la que llevar el control de las digitalizaciones realizadas. Esta base de datos contiene los siguientes campos:

- Número de identificación autonómico.
- Título
- ISSN
- Biblioteca depositaria
- Fecha de publicación
- Volumen
- Número
- Nombre del fichero generado PDF hipervinculado.
- Notas a cada fichero

Próximamente realizaremos una base de datos paralela, relacionada con esta, en la que incluir los correspondientes Metadatos asociados a cada imagen (objetos digitales simples) o a cada grupo unificado de imágenes, cual es el frecuente caso de sumarios que ocupan varias páginas que se escanean independientemente (objetos digitales complejos). El sistema de metadatos que hemos elegido es el RDF (*Resource Description Framework*), estándar creado por el W3C para su presentación en XML.

### 2.3. ESCANEADO

El ordenador de trabajo cuenta con dos discos duros. Uno para el sistema operativo e instalación de las aplicaciones y otro para guardar el trabajo de digitalización. En el disco de trabajo creamos una estructura de directorios en el que cada uno de ellos corresponde con el número de identificación de la publicación correspondiente: 0001, 0002, 0003... y dentro de cada uno de ellos, el año al que pertenece la publicación.

Las imágenes son reconocidas en el escáner mediante el alimentador de páginas, en

blanco y negro a 1 bit y 300 ppp con el software PaperPort. Una vez escaneado un lote de páginas, comprobamos en la pantalla de trabajo si las mismas han sido digitalizadas correctamente sin que estén torcidas, cortadas, la legibilidad del texto, que no tienen márgenes incorrectos, etc.

A continuación seleccionamos todas las páginas correspondientes a un mismo sumario, una o varias, y las convertimos automáticamente a PDF pulsando el botón de conversión que aparece en la misma pantalla principal de PaperPort. El programa Adobe Acrobat se encuentra ajustado previamente para formato A4 y 300 ppp. Una vez convertido a PDF se nos abre la aplicación Acrobat, procediendo nosotros a editar los datos del documento introduciendo los siguientes: en *Título*, el de la publicación; en *Asunto*, el año, volumen y número; en *Autor*, las siglas de la Biblioteca Universitaria (BULPGC). Acto seguido, nombramos el archivo generado para guardarlo en formato "8.3" con la siguiente disposición: *Número identificación revista número sumario.pdf*, como por ejemplo: 0001\_001.pdf. Este fichero es almacenado en su directorio correspondiente identificado, como ya hemos explicado, por el mismo número de identificación asignado al título de la publicación seriada. Acto seguido anotamos en la base de datos de control los números que hemos escaneado.

Al final de cada jornada de trabajo se hace copia de seguridad de todos los datos con la regrabadora de CD-ROM instalada y se transfiere, vía FTP, los ficheros digitalizados al servidor para su actualización en la Web de la BULPGC.

### 3. ACCESO PÚBLICO A LOS SUMARIOS

La última fase consiste en actualizar los datos incorporados. El usuario puede acceder a

los sumarios mediante *Áreas de conocimiento*, por *Bibliotecas* o directamente por *Títulos*. En este último caso da la posibilidad tanto de teclear directamente el nombre de la publicación que nos interesa o su ISSN y también, pulsando bajo la inicial, aparece el listado correspondiente a dicha letra.

Nos encontramos también en fase de desarrollo de un servicio de alerta de nuevos sumarios, por el cual el usuario que se da de alta en el mismo recibe periódicamente en su cuenta de correo electrónico un mensaje comunicándole que ya tiene a su disposición determinados sumarios de su interés con el hipervínculo a los mismos para su consulta directa.

