

Nuevo formato electrónico para la distribución en línea de imágenes documentales de alta calidad

Victor M. Macías Alemán

Biblioteca Universitaria
Universidad de Las Palmas de Gran Canaria

BIBLID 1133-9756 (2001-2002)13-14, p. 17-22

Quienes nos dedicamos, cada vez en mayor medida, a la digitalización documental con el objeto de ofrecer a los usuarios de forma remota un número cada vez mayor de documentación a texto completo en el ámbito de las bibliotecas, hemos recurrido habitualmente a la utilización masiva de formatos extendidos y visualizables directamente en navegadores Web como el *JFIF* (más conocido, aunque erróneamente, por *JPEG*) o a formatos propietarios como PDF, visualizables también en el mismo entorno a través de plug-ins.

Sin embargo, la falta de un manejo adecuado de la imagen –con la necesidad de diferentes ficheros de la misma imagen– para su correcta visualización en el navegador en el caso del primero, y el hecho del gran tamaño de los ficheros resultantes –particularmente si empleamos color o escala de grises– en el caso del segundo, está suponiendo significativos problemas a la hora de hacer eficaz tanto la generación como la distribución de la información electrónica facsímil de documentos escaneados a través de la red.

Ante ello, y pensando en el proyecto de Biblioteca Digital puesto en marcha en el seno de la Biblioteca de la Universidad de Las Palmas de Gran Canaria, nos pusimos a investigar sobre la existencia de otras soluciones que solucionaran dichos problemas. Nuestros esfuerzos se vieron recompensados al conocer la existencia del formato de fichero conocido como DjVu.

UNA SOLUCIÓN PLAUSIBLE: DJVU

DjVu (se pronuncia como las palabras francesas *déjà vu*) es un nuevo –aunque sus orígenes se remontan ya a los años ochenta– formato para la distribución de imágenes facsímiles escaneadas, bien sea mediante su almacenamiento en soportes físicos digitales en disco óptico como a través de Internet. DjVu permite el manejo de imágenes documentales en blanco y negro, grises o color, aspectos estos últimos en los que destaca sobresalientemente ya que, como veremos a continuación, no existe ningún otro formato capaz de transmitir imágenes de estas características con alta calidad y muy escaso tamaño.

Creado por la multinacional estadounidense AT&T, este formato está especialmente indicado para la distribución de imágenes de bibliotecas, archivos, servicios de publicaciones o documentación de entidades comerciales a través de red e Internet, ya que a diferencia de otros formatos, consume una escasa cantidad de ancho de banda en su transmisión proveedor-usuario. Esta ventaja se basa en un sistema multicapa, que logra la reducción del tamaño de los archivos mediante la separación de los documentos en dos capas: alto contraste para el texto (comúnmente los colores oscuros o el negro) y bajo contraste para lo que serían los fondos grises o color propios de, por ejemplo, fotografías o mapas.

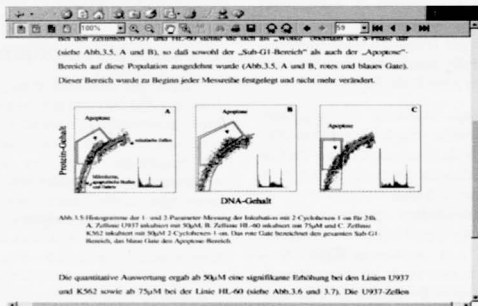
Gracias a su sistema algorítmico, cada capa es codificada de forma que proporcione los mejores resultados para conseguir un escaso tamaño y una buena imagen. Este sistema de estructurar la imagen se aprecia fácilmente en el propio navegador web, ya que la barra de herramientas de navegación nos dá la posibilidad de elegir –poniendo por caso que nos

encontremos en un documento textual con gráficos en color– el ver sólomente el texto, el fondo o ambas a la vez (opción por defecto).

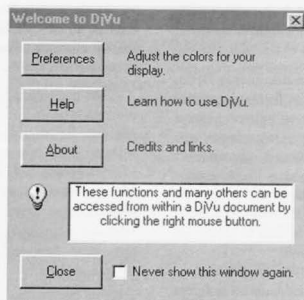
VISUALIZACIÓN Y MANEJO DE DOCUMENTOS DJVU

Existe una amplia gama de plug-ins para la visualización de este formato en los navegadores de uso más común en entornos Windows 95, 98, 2000, ME, NT, o XP (para Internet Explorer y Netscape Navigator), Macintosh, Unix Solaris y Linux.

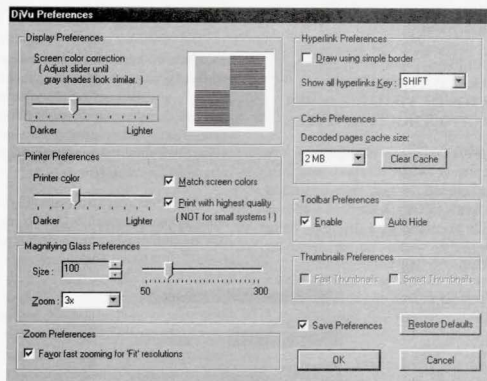
En cuanto a los idiomas de las interfaces, al momento de escribir estas líneas se encuentran solamente en inglés, francés y alemán. En cualquier caso y hasta la aparición de la versión que contemple nuestro idioma, la facilidad de uso de su interface no comporta ningún problema en su uso. Además de ello, pulsando en cualquier lugar de la imagen con el botón derecho del ratón, accederemos a un menú donde podemos seleccionar la opción *Help* (Ayuda)



Aspecto del navegador web DjVu bajo Internet Explorer 5.5 con una tesis doctoral de la Universitätsbibliothek Kaiserslautern



Pantalla de bienvenida al iniciar DjVu



Ventana de configuración de nuestro navegador DjVu



Detalle de la barra de navegación de documentos DjVu

que incorpora gran cantidad de documentación explicando cada aspecto del uso de este formato con nuestro navegador.

La visualización de un documento DjVu, al igual que sucede con otros formatos propietarios, pasa por descargar en primer lugar a nuestro equipo el correspondiente *plug-in* que usualmente estará disponible en el mismo lugar desde el cual nos descarguemos los documentos DjVu. Una vez instalado en nuestro equipo bastará con hacer doble click sobre el fichero descargado o bien sobre el enlace al mismo en la correspondiente página web.

Cuando se visualice el documento se mostrará una ventana como la indicada en la figura siguiente, que nos permitirá parametrizar la visualización que deseemos o acceder a la ayuda, pudiendo también desactivarla para futuras visualizaciones.

COMPARATIVA CON OTROS FORMATOS DE VISUALIZACIÓN DE IMÁGENES DOCUMENTALES

Gracias a los poderosos algoritmos y a la estructura empleada en este formato, para imágenes de alta calidad, la tasa de compresión puede ser de hasta 1.000 a 1. Debido a ello, el ancho de banda necesario para descargar de la red documentos en color no necesita ser

un simple modem de 56 Kb. Comparativamente entre DjVu y PDF, las descargas documentales con el primero pueden llegar a ser hasta 150 veces más rápidas.

Un buen ejemplo de lo indicado sería el escaneado de un documento a todo color de 110 páginas. El original ocupa 2,5 Gigabytes en formato TIFF sin compresión. Convertido a PDF ocuparía 155 Mb, en JFIF (JPEG) 128 Mb, mientras que en DjVu ocupará tan sólo 3 Mb.

Como puede apreciarse la diferencia en espacio, y por lo tanto en tiempo de descarga, es abrumadora. Con respecto a ello, hay que tener en cuenta que el formato PDF fue diseñado fundamentalmente para documentos producidos de forma electrónica, mientras que DjVu fue diseñado para documentos digitalizados mediante escáners.

Por otro lado, otra gran ventaja a la hora de utilizar DjVu es que incorpora tecnología "streaming", por lo que a diferencia del formato PDF no es necesario descargar la totalidad del fichero para poder visualizarlo, sino que, según vayamos leyendo y avanzando páginas, se irán cargando las restantes.

DjVu incorpora asimismo la posibilidad de aplicar reconocimiento óptico de caracteres (OCR) en el mismo modo de "imagen original

155MB PDF



128MB JPG



3MB DjVu



amplio, como sucedería con documentos existentes en formato PDF desarrollado por Adobe, su rival más directo, sino que bastaría con una conexión reducida como la proporcionada por

con texto oculto" incorporada por PDF, con el cual poder extraer y copiar el texto contenido en los documentos digitalizados, a fin de incorporarlo a ulteriores trabajos. En este aspecto, al

igual que en todas las posibilidades que incorpora el visor web –rotación de la imagen, aumento y disminución de la visualización desde 5% a 1200 %, impresión, descarga del documento a disco, etc.– DjVu es plenamente comparable con el formato PDF.

Con respecto a su comparación con el formato JFIF, la compresión JPEG sacrifica la calidad de la imagen, mientras que en DjVu se conservan las características visuales del documento original. Además, para documentos en color a 300dpi que contienen texto y fotografías, los archivos DjVu son diez veces más pequeños que JFIF con calidad similar. Además de esto es conocida la inexistencia de buenos *plug-ins* gratuitos para el manejo de estas imágenes desde dentro de navegadores web, debido fundamentalmente al hecho de ser formato nativo directamente reconocido por los mismos. Por ello, no es posible aumentar o disminuir directamente dichas imágenes, aplicarles OCR, rotarlas, etc. sin que tengamos que crear un sinnúmero de subimágenes JPEG que cubran todas las posibilidades que podría exigir un usuario, aumentando por tanto la cantidad de espacio necesario en el servidor correspondiente para ofrecer la misma información.

UNA AMPLIA GAMA DE SOFTWARE

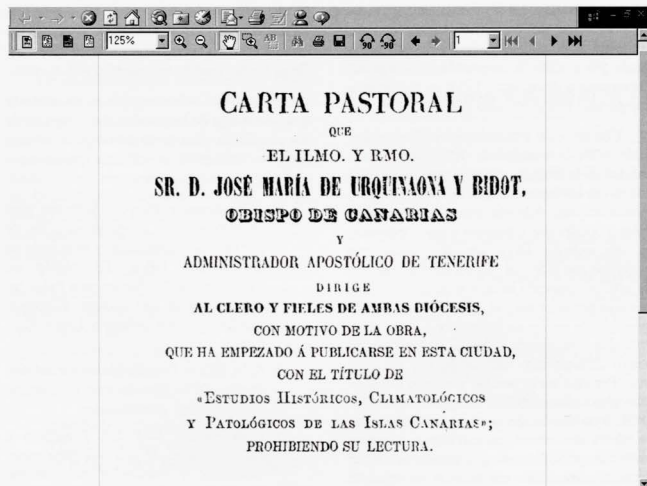
En su versión original de AT&T, DjVu es un formato de libre distribución para el cual se dan todo tipo de facilidades para que particulares y organismos puedan facilitar su desarrollo y ampliación. Por ello se proporciona el código fuente del mismo, además de licencias gratuitas de utilización para propósitos personales y no comerciales particularmente en entorno Linux.

Existen además servicios en línea gratuitos, como *Any2DjVu*, que permiten convertir documentos escaneados en formato JFIF(JPEG), Postscript, PDF, Tiff y otros a

DjVu. Ello se puede realizar bien subiendo dichos documentos al servidor o bien indicando una URL en la que se encuentren ubicados. En el mismo sentido trabaja *Bib2Web Server*.

LizardTech es la compañía estadounidense que se encarga de la distribución comercial de los programas para crear documentos en este formato. La familia de software que ofrece a este respecto es la siguiente:

- *DjVu Browser Plug-in*. Software para incorporar a los navegadores a fin de permitir la visualización y el manejo de documentos DjVu. Disponible en inglés, francés y alemán para todas las plataformas de uso común: Windows, Macintosh, Unix-Solaris y Linux. Gratuito.
- *DjVu Editor*. Permite añadir a los documentos DjVu hiperenlaces y anotaciones entre otras posibilidades.
- *DjVu Encode SDK 3.5*. Destinado a desarrolladores de software para crear, visualizar y manipular documentos Djvu nativos a fin de crear aplicaciones por terceras partes.
- *DjVu Enterprise 3.5*. Permite la conversión masiva de grandes cantidades de documentación a DjVu.
- *DjVu Workgroup 3.0*. Similar al programa ya mencionado Enterprise 3.5, permite la conversión masiva de documentación con opciones como arrastrar y soltar.
- *DjVu Solo 3.1*. Permite la conversión de páginas simples, una a una, de documentación electrónica a formato DjVu. Se ofrece en dos versiones, edición personal que no incluye OCR y es gratuita y la edición comercial que sí incluye esta característica.



Documento DjVu creado con el software gratuito DjVu Solo 3.1

ENLACES

Any2DjVu. Conversión en línea gratuita a DjVu. <http://any2djvu.djvuzone.org/>

Bib2Web. Opera de forma similar al anterior. <http://bib2web.djvuzone.org/>

DjVu Zone (AT&T). Acceso al código

fuente y programas gratuitos para Linux <http://www.djvu.att.com>

LizardTech. Dedicado a la distribución de software comercial en entorno DjVu así como de la versión gratuita de conversión DjVu Solo. Incorpora enlace a múltiples ejemplos. <http://www.lizardtech.com/products/djvu.php>