La compresión en el formato PDF

Laurens Leurs

(Abril de 2001)

Los ficheros PDF son mucho menores que sus equivalentes PostScript. Aunque esto se consigue en parte gracias a una mejor estructura de los ficheros, la razón principal se debe a los algoritmos de compresión extremadamente eficaces que admiten los PDFs.

Todos los textos y todas las órdenes PostScript que forman parte de un PDF se pueden comprimir usando el algoritmo LZW (Lempel-Ziv-Welch). Esta compresion por si sola ya puede dejar el tamaño de un fichero PDF en la mitad del tamaño de su equivalente PostScript.

Si la comresión está activada en el Acrobat Distiller, lo primero que hará este programa es descomprimir todas las imágenes que se hallen dentro del fichero PostScript y luego las volvera a comprimir al crear el fichero PDF. En el caso de las imágenes, existen varios algoritmos de compresión admisibles: JPEG, Zip, CCITT, RLE y "hinchar/deshinchar" *(flate)*.

Para comprobar qué tipo de compresión se ha utilizado, basta con abrir el fichero PDF en un simple editor de texto que admita datos binarios (el Bloc de Notas, etc...) y buscar allí la cadena "`/Filter`".

La compresión JPEG

La compresión JPEG se usa para imágenes de cuatricromía y escala de grises. Es un algoritmo de compresión que puede ser "con pérdidas" *(lossy)* y "sin pérdidas" *(non-lossy).* Sin embargo, en Adobe Acrobat sólo está disponible el modelo "con pérdidas". Esto quiere decir que algunos detalles de la imagen se pierden o deterioran cada vez que se la comprime. Cuanto mayor es la tasa de compresión, mayor es el deterioro de las imágenes. Si quieres saber más sobre el algoritmo de compresión JPEG, puedes leer esto [y también esto otro, en inglés].

Acrobat Distiller 3.x ofrecía cinco niveles de compresión JPEG:

1. Baja: Propociona una pérdida de calidad que es generalmente aceptable salvo en el caso de los trabajos más exigentes. La tasa media de compresión es de 1:2.
2. Baja - media
3. Media: Proporciona una tasa de calidad para trabajos de tipo medio. La tasa media de compresión es de 1:5.
4. Alta - media
5. Alta: not acceptable for prepress any more. Average compression ratio: 1/10

Acrobat Distiller 4 tenía disponibles ajustes muy similares, pero no se correspondían exactamente con los de Acrobat 3. En Acrobat Distiller 4.0 había además un fallo de programación que afectaba al algoritmo JPEG. Este fallo se arregló con la actualización a 4.05. Las opciones son:

1. Mínima.
2. Baja.
3. Media.
4. Alta.
5. Máxima.

Las imágenes que se han comprimido usando el algoritmo JPEG no deben recomprimirse, ya que esto causará una pérdida acumulativa de calidad. Si se destila un fichero que contiene imágenes comprimidas como JPEGs, Acrobat Distiller las descomprimirá y volverá a comprimir (dependiendo de los parámetros que se hayan establecido). Esto dará lugar a la mencionada pérdida de calidad.

La extensión para Adobe Acrobat Quite A Box Of Tricks (de Quite Software) permite algoritmos de compresión JPEG más potentes que los del propio Acrobat. Además permite la recompresión de ficheros PDF.

La compresión Zip

El algoritmo de compresión Zip es el que usan utilidades para PC muy populares como WinZip, PKZip, Stuffit y similares. su uso no quiere decir que Acrobat sea capaz de crear un fichero Zip, sino que usará el algoritmo para comprimir imágenes de color o escala de grises.

La compresión Zip es un poco más "inteligente" que la realizada por el algoritmo LZW. Lo que hace es escanear un fichero para detectar secuencias repetidas y entonces reemplaza éstas por una marca. Un ejemplo sería comprimir todo la *web* del Rincón del PDF y sustituir la palabra "fichero" (de gran frecuencia) por

una simple marca. Una sucesión de sustituciones permite un gran ahorro de bytes.

Zip es un algoritmo de compresión "sin pérdidas". Esto quiere decir que las imágenes no se degradarán al comprimirlas. En Adobe Acrobat 4 es realmente así, pero en Acrobat 3 existe la opción de usar una compresión de 4 bits y otra de 8 bits. El problema es que con la de 4 bits, Acrobat reducirá primero el número de tonos por canal de 256 a sólo 16 y, posteriormente, realizará la compresión "sin pérdidas". El resultado es una tasa de compresión excelente y una degradación de la calidad extrema. Por consiguiente, la compresión Zip de 4 bits se debe evitar siempre a no ser que las características (color, contenido, etc...) de los ficheros se presten a ello. La compresión zip con 8 bits carece absolutamente de pérdidas.

La compresión CCITT

La compresión CCITT se puede usar con imágenes de línea (en Photoshop: Mapa de bits). Es el mismo algoritmo de compresión que utilizan los faxes. Es una compresión "sin pérdidas", por lo que las imágenes no sufren degradación alguna.

Adobe Acrobat ofrece la posibilidad de elegir entre Grupo 3 de CCITT y Grupo 4 de CCITT. La mayoría de las opiniones están de acuerdo en que la segunda es mejor. Se puede dejar activada por omisión.

La compresión RLE

El algoritmo RLE (Run Length Encoding) es un sistema de compresión "sin pérdidas", por lo que no afectará a la calidad de las imágenes. El algoritmo RLE es probablemente el más simple que existe. en vez de escribir un fichero que vaya diciendo: "Pixel blanco, pixel negro, pixel negro, ..." Va diciendo algo así como: "129 píxeles negros, dos píxeles blancos, ..."

En Acrobat, la compresión RLE se puede usar para imágenes de línea (en Photoshop: Mapa de bits). La mayoría de los usuarios prefieren la compresión CCITT porque es más eficaz.

La compresión "hinchar/deshinchar" (flate)

El algoritmo de "hinchar/deshinchar" (*flate,* a veces también llamada *deflate*) es un sistema de compresión bastante complejo que he tratado con más

profundidad en otra página.