

Herramienta Multifuncional para ficheros PDF

PDF A TOPE

Para organizar la montaña de papeles que se cruzan cada día en nuestra mesa necesitamos organizar, recuperar, copiar, imprimir, buscar y clasificar documentos. Una herramienta especial puede ayudar a los usuarios a interesarse por el trabajo administrativo electrónico: pdftk, el juego de herramientas PDF.

POR STEFAN LAGOTZKI

Utilidades PDF nativas de Linux, tales como GhostScript, son muy útiles si estás dispuesto a navegar a través de los menús. Pero si estás buscando algo más rápido, o si lo que prefieres es automatizar alguna tarea recurrente, prueba con pdftk, el juego de herramientas PDF. pdftk es un programa de la línea de comandos muy apropiado para procesar ficheros PDF. Según su creador, Sid Steward, "Si PDF es papel electrónico, entonces pdftk es desgrapadora electrónica, taladrador, encuadernador y gafas de rayos X electrónica.

Instalación y Uso

Puedes descargar la última versión del juego de herramientas de una de las páginas web de Sid Steward [1]. El programa GPL se encuentra disponible para Linux, Mac OS X (Panther), FreeBSD, Solaris y Windows. La instalación resultó ser bastante simple sobre las plataformas que probamos (Debian y SuSE Linux).

Tras completar la instalación puedes correr pdftk desde la shell. El comando de *-help* de pdftk ofrece una lista de comandos y opciones con textos de ayuda cortos. La tabla 1 lista y explica las operaciones más importantes. La sin-

taxis genérica para el procesamiento de ficheros PDF con el programa es:

```
pdftk fichero_entrada >
operación [opción] >
output fichero_salida >
[contraseñas] [permisos_usuarios]
```

Los ficheros de entrada han de estar en formato PDF. Además, la herramienta necesita ficheros de texto en un formato especial para algunas operaciones. pdftk da como salida uno o más ficheros PDF así como ficheros de texto en algunos casos especiales.

En las secciones que siguen he dispuesto unos cuantos ejemplos que demuestran un poco los usos más interesantes de pdftk. Estos ejemplos no hacen más que rayar la superficie de las posibilidades de este complejo programa.

Adjuntos en ficheros PDF

Se puede añadir un adjunto a un fichero PDF del mismo modo que puedes añadirlo a un correo electrónico. Adobe Reader (versión 6 o superior) almacena los adjuntos en el recipiente. pdftk permite a sus usuarios adjuntar ficheros a documentos PDF y también guardarlos.

Antes de la publicación de Adobe Reader 7, era la única manera de que los usuarios Linux pudieran guardar adjuntos.

Los adjuntos pueden ser usados para añadir código fuente o extractos de base de datos literarias a ficheros PDF. El siguiente ejemplo muestra cómo enviar un fichero PDF y el código fuente acompañante. Para añadir un fichero de código fuente a un documento PDF

Listado 1: Ejemplo de Fichero PDF

```
01 %PDF-1.2
02 1 0 obj <<
03 /FDF << /Fields [
04 << /V (Dresden)/T (city) >>
05 << /V (Stefan Lagotzki)/T
06 (author)>>
07 ]/F (form.pdf) >>
08 >>
09 endobj
10 trailer
11 <<
12 /Root 1 0 R
13 >>
14 %%EOF
```

usando `pdftk`, escribe los siguientes comandos:

```
pdftk fichero.pdf attach_files >
fichero.text output nuevo.pdf
```

Alternativamente, podrías usar `pdfLaTeX` y un fichero adjunto para añadir el código fuente al fichero PDF acabado. El destinatario usaría luego `pdftk` para desempaquetar el fichero código fuente y otros adjuntos usados en cualquier directorio:

```
pdftk exejemplo_adjunto.pdf >
unpack_files output Fuente
```

En este ejemplo, guarda los adjuntos en un directorio llamado *Fuente*. Añadir el nombre del directorio siempre tiene sentido si estás manejando múltiples adjuntos.

Marcas de Agua y Colores

`pdftk` utiliza un método similar al paquete LaTeX *eso-pic* para añadir una marca de agua a un documento. La opción *background* es la que maneja esto, además de permitir al usuario asignar un color a la imagen de fondo PDF.

La imagen que uses como marca de agua deberá ser un fichero PDF. Podrías crear una imagen con una herramienta de gráficos vectorial o bien escribir un programa PostScript. Si la marca de agua no es del mismo tamaño que el documento, `pdftk` escalará la marca de agua. Asigna a lo que quieres imprimir la palabra "DRAFT" en un documento. El primer paso sería crear un PDF con el tamaño de la página adecuado antes de llamar a `pdftk` como sigue:

```
pdftk ejemplo.pdf background >
borrador.pdf output
borrador1.pdf
```

La marca de agua se asemeja a un sello en cualquier parte del documento sin contenido. Puedes crear un pequeño fichero EPS con un color de imagen de fondo a tu elección. Los comandos PostScript para una página de tamaño A4 se parecen a:

```
%!PS-Adobe-2.0
%%BoundingBox: 0 0 595 842
0.95 0.95 0.90 setrgbcolor
0 0 moveto 595 0 rlineto 0 842 >
rlineto -595 0 rlineto
```

```
closepath fill
showpage
```

Es fácil cambiar el color de la imagen de fondo en el código EPS. Puedes entonces convertir el fichero EPS a PDF usando *epstopdf* y correr luego `pdftk` para usar el fichero como una imagen de fondo:

```
pdftk ejemplo.pdf background >
Fondo.pdf output eg_color.pdf
```

Dividiendo y Uniendo Ficheros PDF

La operación *burst* te permite dividir un fichero PDF en sus páginas individuales. Para hacerlo, necesitas suministrar un nombre genérico para las páginas y especificar el formato numérico:

```
pdftk ejemplo.pdf burst >
output Pagina%03d.pdf
pdftk ejemplo.pdf burst >
output ./Paginas/Paigna%03d.pdf
```

En ambos ejemplos, un número de página de tres dígitos se añadirá a los nombres de la página. En el segundo ejemplo, `pdftk` almacenará los ficheros PDF en un subdirectorio existente.

La operación *cat* le dice a `pdftk` que concatene múltiples ficheros PDF para crear un nuevo documento. Puedes usar comodines para especificar los nombres de ficheros de los ficheros fuente individuales.

```
pdftk ejemplo.pdf otro.pdf >
adjunto.pdf >
cat output ejemplo_concat.pdf
pdftk D=cubierta.pdf >
```

```
B=ejemplo.pdf >
cat D B1-4 output >
ejemplo_cubierta.pdf
```

Tal y como demuestra el segundo ejemplo, puedes usar *cat* para reordenar documentos mediante la unión de las partes de un fichero PDF con partes de otro fichero PDF para crear un nuevo documento.

Consulta y Actualización de Meta-Información

La mayoría de los ficheros PDF contienen meta-información con detalles del autor, el tema o el software usado para crear el documento. `pdftk` te permite enviar estos datos a una salida estándar o redireccionarlos a un fichero:

```
pdftk ejemplo.pdf >
dump_data output info.txt
```

Este comando salva la información desde un documento PDF a un fichero titulado *info.txt*. La información incluye un campo clave y su valor. Antes de enviar o de archivar documentos PDF, a menudo tiene sentido actualizar los meta-datos. `pdftk` te permite hacerlo sin tener que recrear o trasladar el fichero PDF.

Para actualizar la meta-información, crea primero un fichero de texto con los meta-datos. El fichero debería ser similar a algo como lo que sigue (resumido por aquello de la brevedad):

```
InfoKey: Creador
InfoValue: TeX
InfoKey: Corporación
InfoValue: Ejemplo e Hijos
```

TABLA 1: OPERACIONES PDFTK

Operación	Explicación
attach_files	Añade ficheros como adjuntos a un documento PDF. Esto permite añadir un fichero de archivo a un fichero PDF.
background	Añade una marca de agua a cada página para un documento PDF. También permite añadir un sello electrónico a espacios vacíos.
burst	Divide un documento PDF en páginas individuales.
cat	Concatena un nuevo fichero PDF a partir de ficheros múltiples o páginas desde documentos PDF distintos.
dump_data	Salida de información sobre un fichero PDF o salida estándar.
dump_data_fields	Salidas de información sobre los campos de formularios en un fichero PDF o salida estándar.
fill_form	Llena los formularios PDF o datos en forma de enlaces con el documento.
unpack_files	Desempaqueta los adjuntos de un documento PDF en un directorio.
update_info	Actualiza la meta-información (por ej. autor, título, tema) en un fichero PDF.

TABLA 2: PERMISOS PDF

Opción	Explicación
Printing	El documento puede ser impreso en la mejor calidad.
DegradedPrinting	El documento a imprimir será de calidad limitada.
ModifyContents	El documento que contiene puede ser modificado.
Assembly	El documento PDF puede ser concatenado con otros documentos PDF.
CopyContents	Texto e imágenes pueden ser copiadas desde el fichero.
ModifyAnnotations	Comentarios y anotaciones pueden cambiarse.
FillIn	Formularios en el fichero PDF pueden ser completados.
AllFeatures	El usuario tiene todos los privilegios especificados.

No necesita contener toda la información que un fichero PDF puede almacenar.

Los campos que ya contienen valores no se tocan en la actualización si el fichero de texto no lo especifica. Puedes incluso añadir nuevos campos clave (Corporación en nuestro ejemplo) y asignarles valores. La siguiente llamada actualiza la información:

```
pdfTk ejemplo.pdf >
update_info info.txt >
output eg_meta.pdf
```

Los ficheros de entrada y salida no tienen permitido usar el mismo nombre. En otras palabras, necesitas renombrar manualmente el fichero de salida o usar un script de la shell para hacerlo.

Cumplimentado de Formularios PDF

Los ficheros PDF pueden contener formularios con campos de formulario conocidos. Adobe desarrolló el formato FDF, que es propietario pero abierto, para datos en formulario PDF. El Listado 1 muestra un ejemplo de un fichero FDF resumido.

En el Listado 2, *T* es el título, y *V* es el valor del campo del formulario. Ahora puedes unir el fichero PDF con el FDF y decidir si los datos del formulario permanecerán editables o estarán unidos indeleblemente con el documento:

```
pdfTk form.pdf fill_form >
eg.fdf output edit.pdf
pdfTk form.pdf fill_form >
eg.fdf output end.pdf flatten
```

El primer ejemplo ofrece los resultados editables. Sin embargo, la opción *flatten* en el segundo fichero indica que los campos del formulario deberían estar unidos indeleblemente con el fichero PDF.

Las características del formulario te permiten usar pdfTk para crear formula-

rios PDF completados sobre un servidor Intranet o Internet. El usuario completa los campos del formulario en su navegador. Luego un script PHP o Perl corriendo de fondo crea un fichero FDF. Finalmente, pdfTk combina ambas partes. El fichero PDF completado puede ser enviado por correo luego.

Contraseñas y Permisos de Usuario en PDFs

Los ficheros PDF pueden estar protegidos mediante el permiso de usuario y contraseñas. pdfTk permite establecer para un fichero PDF tanto las contraseñas como los permisos. El siguiente ejemplo establece ambas contraseñas:

```
pdfTk fichero.pdf output >
nuevo.pdf owner_pw >
Lie5quai user_pw phupaefu
```

En este ejemplo las contraseñas fueron generadas usando la herramienta *pwgen*. Deberías usar distintas para las contraseñas y permisos.

El propietario de un fichero PDF puede asignar permisos específicos. La Tabla 2 presenta una lista de permisos que pueden establecerse con pdfTk.

El siguiente ejemplo crea primero un fichero PDF que solamente puede imprimirse. La segunda línea crea un fichero PDF que además de imprimirse puede ser también copiado.

```
pdfTk ejemplo.pdf output >
nuevo.pdf owner_pw >
Lie5quai user_pw phupaefu >
allow printing
pdfTk ejemplo.pdf output >
nuevo.pdf owner_pw >
Lie5quai user_pw phupaefu >
allow printing CopyContents
```

Los ficheros PDF pueden ser encriptados con diferentes niveles de encriptación. Para encriptar un fichero con pdfTk hay

que añadir como opción final *encrypt_40bit* o *encrypt_128bit*. También puedes necesitar sustituir una contraseña por un fichero PDF protegido con contraseña. Si estás procesando múltiples ficheros, puedes imponer variables a los nombres de ficheros y luego asignar una contraseña a cada fichero. En el ejemplo siguiente, solamente el fichero A tiene protegida la contraseña:

```
pdfTk A=nuevo.pdf >
B=eg_color.pdf input_pw >
A=Lie5quai cat output >
eg1_pw.pdf user_pw Abraxas
```

Como en el ejemplo anterior no te permite concatenar el fichero PDF por *nuevo.pdf*, necesitas sustituir la contraseña del propietario.

Conclusiones

Si estás buscando una herramienta simple, rápida y eficiente para editar ficheros PDF desde la línea de comandos, prueba con el Toolkit PDF. pdfTk es una herramienta de manipulación PDF versátil y multifuncional sin el peso de un GUI. Si quieres hurgar profundamente en la manipulación de ficheros PDF, échale un vistazo al libro de Sid Steward sobre *Hacks de PDF* [2].

pdfTk está escrito en C++ y basado en la librería iText [3], escrita en Java. El programa completo fue compilado y linkado con herramientas libres como el GNU Compiler Collection [4]. Esto hace a pdfTk fácilmente portable y extensible. La web-site de pdfTk tiene enlaces a portes.

El trabajo de desarrollo sobre el programa pdfTk aún continua. Su autor, Sid Steward responderá a todas las cuestiones sobre pdfTk y correos sobre programación PDF en el grupo de noticias *comp.text.pdf* y en su propio foro PDF [1].

RECURSOS

- [1] Sid Steward: pdfTk; Versión 1.12 (Nov. 2004): <http://www.accesspdf.com/pdfTk/>
- [2] Sid Steward, *PDF Hacks*, O'Reilly, 2004.
- [3] Bruno Lowagie, Paulo Soares: iText-Library; Versión 1.1 (Nov. 2004): <http://itext.sourceforge.net>
- [4] GNU Compiler Collection, Versión 3.4.3 (Nov. 2004): <http://gcc.gnu.org>