

Concatenar pdf

2016-05-11

Introducción:

El objetivo es concatenar múltiples pdf en un único pdf, intentando mantener la estructura interna de los pdf a unir.

Para hacer esto tenemos múltiples opciones:

- Concatenar por línea de comandos:
 - pdffunite
 - ImageMagick
 - pdffjam
 - Ghostscript Pad
- Concatenar con librerías externas:
 - Python
- Concatenar Online:
 - smallpdf
- Resumen general
- Conclusión final

Concatenar por línea de comandos

- pdffunite
- ImageMagick
- pdffjam
- Ghostscript Pad

Pdffunite

Pdffunite es una herramienta muy simple escrita en ruby para unir pdf, se instala rápido, funciona bien pero el índice de los pdf se lo carga.

Instalación.

Solo hay que instalar el paquete `poppler` que lo contiene.

```
dnf install poppler-utils
```

Unión de pdf.

Para unir varios pdf lo único que hay que hacer es pasárselos a pdffunite como argumento teniendo en cuenta que se irán añadiendo al pdf en el mismo orden y que el último parámetro debe de ser el nombre del pdf resultante.

```
pdffunite pdf1.pdf pdf2.pdf output.pdf
```

Como inconveniente el índice que traigan otros pdf se lo carga.

ImageMagick (no recomendado)

Este método funciona pero reduce la calidad (resolución) del pdf un poco y aumenta bastante el tamaño del pdf resultante, su uso sería el siguiente:

```
convert -density 300x300 -quality 100 pdf1.pdf pdf2.pdf output.pdf
```

Pdfjam

Pdfjam es un conjunto de pequeños script para hacer modificaciones a pdf, pero a nosotros la única herramienta que nos interesa ahora mismo es pdfjam que es la principal, ya que es la que usaremos para unir los pdf.

Instalación.

La Instalación en Fedora es simple, solo hay que instalar un paquete.

```
[sudo] dnf install pdfjam
```

Unión de pdf.

Para unir varios pdf lo único que hay que hacer es pasárselos a pdfjam como argumento teniendo en cuenta que se irán añadiendo al pdf en el mismo orden y que el pdf resultante debe ir precedido por un `-o` para indicar que es la salida.

```
pdfjam pdf1.pdf pdf2.pdf -o output.pdf
```

Como inconveniente el índice que traigan otros pdf se lo carga.

Ghostscript

Ghostscript es un paquete que te permite ver o imprimir archivos PostScript y PDF en otros formatos o a partir de estos a otros formatos, pero a demás de todo eso también te permite unir archivos pdf. La instalación no es necesaria al menos en Fedora 24 ya que venia instalado por defecto.

Unión de pdf.

En este apartado al ser una herramienta bastante completa podríamos hacerlo de múltiples formas, antes de ponerla pondré algunos parámetros importantes que pasaremos a Ghostscript:

- `-dBATCH`: una vez que Ghostscript procesa los archivos PDF, debe salir. Si no incluye esta opción, Ghostscript seguirá ejecutándose.
- `-dNOPAUSE`: obliga a Ghostscript a procesar cada página sin pausar la interacción del usuario.
- `-q`: para que Ghostscript no muestre mensajes mientras funciona.
- `-sDEVICE=pdfwrite`: le dice a Ghostscript que use su escritor de PDF incorporado para procesar los archivos.
- `-sOutputFile=output.pdf`: le dice a Ghostscript que guarde el archivo PDF combinado con el nombre que específico.

Unir varios pdf.

```
gs -dBATCH -dNOPAUSE -q -sDEVICE=pdfwrite -sOutputFile=output.pdf pdf1.pdf pdf2.pdf
```

Unir varios pdf haciendo una pequeña compresión.

```
gs -dBATCH -dNOPAUSE -q -sDEVICE=pdfwrite -dPDFSETTINGS=/prepress -sOutputFile=output.pdf pdf1.pdf pdf2.pdf
```

Unir varios pdf haciendo una gran compresión, revisar porque puede haber perdidas de calidad visibles.

```
gs -dBATCH -dNOPAUSE -q -sDEVICE=pdfwrite -dPDFSETTINGS=/default -dCompatibilityLevel=1.4 -dQUIET -dDetectDup
```

Fuentes: [stackoverflow](#), [linux.com](#)

Concatenar con librerías externas

- Python

Python

En python podemos usar la librería pyPdf2 que depende de la librería pdf. El resultado es un pdf con el nombre `salida.pdf` que tiene buena calidad pero no mantiene los indices

IMPORTANTE: Actualizar el script para pasarle distintos argumentos como `-o` para hacerlo dinamico y que no tenga un fichero de salida statico

Instalación.

```
[sudo] pip install pdf pyPdf2
```

Unir varios pdf. Uso: `python concatena_pdf.py portada.pdf memoria.pdf`

Descargar script

```
#!/usr/bin/env python
from pyPdf import PdfFileReader, PdfFileWriter
import sys
#Fuente: http://stackoverflow.com/questions/3444645/merge-pdf-files

# Creating a routine that appends files to the output file
def append_pdf(input,output):
    [output.addPage(input.getPage(page_num)) for page_num in range(input.numPages)]

if len(sys.argv) == 3:
    # Creating an object where pdf pages are appended to
    output = PdfFileWriter()

    # Appending two pdf-pages from two different files
    for pdf in sys.argv[1:]:
        append_pdf(PdfFileReader(file(pdf, "rb")),output)

    # Writing all the collected pages to a file
    output.write(open("salida.pdf", "wb"))
```

Concatenar Online

- smallpdf

Smallpdf

Smallpdf es una pagina online en la que podemos hacer múltiples operaciones con pdf como unirlos, separarlos y convertirlos a múltiples formatos, funciona bastante bien pero supongo que tendrá limitaciones en cuanto al tamaño de los ficheros y el numero máximo de operaciones a hacer con una cuenta gratis/sin registrar.

Resumen general

herramienta	mantiene integridad del indice	calidad resultante
pdfunite		
ImageMagick		
pdfjam		
Ghostscript		
pyPdf2		
smallpdf		

Conclusión final

Creo que en este caso la mejor opción es Ghostscript ya que es la única que mantiene el índice después de la unión.

Fuente: 0, 1