Manipulations de fichiers PDF

Page créée le 29 mars 2023

Manipulation de fichiers pdf en ligne de commande, par exemple pour des traitements par lots. Il existe de nombreux utilitaires sous linux dont les fonctions sont parfois identiques. Cette page présente différents exemples d'utilisation qui nous ont été utiles au fil du temps!

Le paquet poppler-utils comprend plusieurs utilitaires dont **pdfseparate, pdffonts, pdfattach, pdfunite**, etc.

pdfjam fait partie du paquet texlive-extra-utils, et permet d'utiliser les fonctions de pdfpages pour LaTeX. «*pdfjam is a shell-script front end to the LaTeX '*pdfpages*' package*»

- https://github.com/pdfjam/pdfjam
- doc pdfpages : https://mirror.ibcp.fr/pub/CTAN/macros/latex/contrib/pdfpages/pdfpages.pdf

```
pdfjam --help
```

ghostscript

- https://manpages.debian.org/bookworm/ghostscript/gs.1.en.html

Autres logiciels utiles

- PDFtk (basé sur java) https://www.pdflabs.com/tools/pdftk-the-pdf-toolkit/
- PDFsam (basé sur java) https://pdfsam.org/ (interface graphique, payant)
- PSPDFUtils (basé sur python) https://pypi.org/project/pspdfutils/
- masterpdfeditor https://code-industry.net/free-pdf-editor/

Extraire des feuillets d'un document pdf

Extraire toutes les pages

```
pdfseparate -f 1 document.pdf pages_%d.pdf
```

Extraire la page 3

```
pdfseparate -f 3 -l 3 document.pdf page3.pdf
```

-f (first) : première page à extraire
-l (last) : dernière page à extraire

Divers

Ajouter un fond perdu

En anglais, fond perdu = *bleed*
La commande suivante ajoute un fond perdu de 3mm sur chaque côté d'une page A5 (148x210mm), un cadre entoure la page originale (`–frame true`), les fichiers *embedded* sont conservés.

```
pdfjam --scale 1.0 --frame true --noautoscale true --papersize '{154mm,216mm}' -o document.pdf document_avec_fond_perdu.pdf
```

–frame true : cadre noir autour de la page originale

Infos sur les polices d'un document

```
pdffonts document.pdf
```

Ces infos permettent de savoir si le fichier de fonte est intégré (*embedded*) dans le fichier pdf, son type, etc. cf. doc pdffonts

Optimiser un pdf pour réduire la taille du fichier (avec ghostscript)

```
gs -sDEVICE=pdfwrite -dCompatibilityLevel=1.4 -dEncodeColorImages=false -dNOPAUSE -dQUIET -dBATCH -sOutputFile=optimized.pdf document.pdf
```

-dEncodeColorImages=false : ne pas réencoder les images JPEG
Si on souhaite qu'aucune modification ne soit appliquée aux images on peut ajouter (source) :

```
  -dColorConversionStrategy=/LeaveColorUnchanged \
  -dEncodeColorImages=false \
  -dEncodeGrayImages=false \
  -dEncodeMonoImages=false \
```

Article extrait de : https://lesporteslogiques.net/wiki/ - **WIKI Les Portes Logiques**
Adresse : **https://lesporteslogiques.net/wiki/ressource/logiciel/manipulation_pdf/start?rev=1737062078**
Article mis à jour: **2025/01/16 22:14**