

Manipuler des PDF sous Linux ?

Liste des programmes

pdfshuffler (PDF file merging, rearranging, and splitting)
Concaténer, séparer ou modifier l'ordre des pages des fichiers PDF.

pdfmod (A simple application for modifying PDF documents)

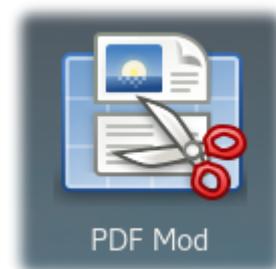
poppler (Command line utilities for converting PDF files)

pdftk

pdf chain

pdfmod

Installer pdfmod



Executer l'installation :

```
sudo dnf install pdfmod
```

Résultat :

```
Dernière vérification de l'expiration des métadonnées effectuée il y a
1:48:37 le sam. 29 juil. 2017 18:27:15 CEST.
Dépendances résolues.
=====
====
Paquet          Architecture  Version          Dépôt
Taille
=====
====
Installation de :
 pdfmod          x86_64        0.9.1-15.fc26    fedora
723 k
Installation des dépendances:
```

gnome-sharp	x86_64	2.24.2-13.fc26	fedora
315 k			
gtk-sharp2	x86_64	2.12.42-2.fc26	fedora
702 k			
hyena	x86_64	0.5-15.fc26	fedora
355 k			
nunit	x86_64	3.6-2.fc26	fedora
369 k			
poppler-sharp	x86_64	0.0.3-13.fc26	fedora
34 k			

Résumé de la transaction

=====
=====

Installer 6 Paquets

Taille totale des téléchargements : 2.4 M

Taille des paquets installés : 8.4 M

Voulez-vous continuer ? [o/N] :o

Téléchargement des paquets :

(1/6): gnome-sharp-2.24.2-13.fc26.x86_64.rpm	190 kB/s	315 kB	00:01
(2/6): hyena-0.5-15.fc26.x86_64.rpm	217 kB/s	355 kB	00:01
(3/6): poppler-sharp-0.0.3-13.fc26.x86_64.rpm	158 kB/s	34 kB	00:00
(4/6): pdfmod-0.9.1-15.fc26.x86_64.rpm	206 kB/s	723 kB	00:03
(5/6): gtk-sharp2-2.12.42-2.fc26.x86_64.rpm	159 kB/s	702 kB	00:04
(6/6): nunit-3.6-2.fc26.x86_64.rpm	333 kB/s	369 kB	00:01

Total	478 kB/s	2.4 MB	00:05
-------	----------	--------	-------

Test de la transaction en cours
La vérification de la transaction a réussi.

Lancement de la transaction de test

Transaction de test réussie.

Exécution de la transaction

- Préparation :
- 1/1 Installation de : gtk-sharp2-2.12.42-2.fc26.x86_64
- 1/6 Exécution du scriptlet: gtk-sharp2-2.12.42-2.fc26.x86_64
- 1/6 Installation de : gnome-sharp-2.24.2-13.fc26.x86_64
- 2/6 Exécution du scriptlet: gnome-sharp-2.24.2-13.fc26.x86_64
- 2/6 Installation de : poppler-sharp-0.0.3-13.fc26.x86_64
- 3/6 Installation de : nunit-3.6-2.fc26.x86_64
- 4/6 Installation de : hyena-0.5-15.fc26.x86_64
- 5/6

```

Installation de      : pdfmod-0.9.1-15.fc26.x86_64
6/6
Exécution du scriptlet: pdfmod-0.9.1-15.fc26.x86_64
6/6
Vérification de     : pdfmod-0.9.1-15.fc26.x86_64
1/6
Vérification de     : gnome-sharp-2.24.2-13.fc26.x86_64
2/6
Vérification de     : gtk-sharp2-2.12.42-2.fc26.x86_64
3/6
Vérification de     : hyena-0.5-15.fc26.x86_64
4/6
Vérification de     : poppler-sharp-0.0.3-13.fc26.x86_64
5/6
Vérification de     : nunit-3.6-2.fc26.x86_64
6/6

Installé :
pdfmod.x86_64 0.9.1-15.fc26          gnome-sharp.x86_64 2.24.2-13.fc26
gtk-sharp2.x86_64 2.12.42-2.fc26    hyena.x86_64 0.5-15.fc26
nunit.x86_64 3.6-2.fc26              poppler-sharp.x86_64 0.0.3-13.fc26

Terminé !

```

pdf shuffler

Installer pdf shuffler



Exécuter l'installation :

```
sudo dnf install pdfshuffler
```

Résultat :

```

Dernière vérification de l'expiration des métadonnées effectuée il y a
1:59:55 le sam. 29 juil. 2017 18:27:15 CEST.
Dépendances résolues.
=====
====
Paquet                Architecture  Version      Dépôt

```

Taille

=====
Installation de :

pdfshuffler	noarch	0.6.0-10.fc26	fedora
-------------	--------	---------------	--------

69 k

Installation des dépendances:

pyPdf	noarch	1.13-13.fc26	fedora
pypoppler	x86_64	0.12.2-5.fc26	fedora

73 k
31 k

Résumé de la transaction

=====
Installer 3 Paquets

Taille totale des téléchargements : 173 k

Taille des paquets installés : 636 k

Voulez-vous continuer ? [o/N] :o

Téléchargement des paquets :

(1/3): pypoppler-0.12.2-5.fc26.x86_64.rpm	19 kB/s		31 kB	00:01
(2/3): pdfshuffler-0.6.0-10.fc26.noarch.rpm	19 kB/s		69 kB	00:03
(3/3): pyPdf-1.13-13.fc26.noarch.rpm	15 kB/s		73 kB	00:04

Total 29 kB/s | 173 kB 00:05

Test de la transaction en cours

La vérification de la transaction a réussi.

Lancement de la transaction de test

Transaction de test réussie.

Exécution de la transaction

Préparation :

1/1
Installation de : pypoppler-0.12.2-5.fc26.x86_64
1/3
Installation de : pyPdf-1.13-13.fc26.noarch
2/3
Installation de : pdfshuffler-0.6.0-10.fc26.noarch
3/3
Exécution du scriptlet: pdfshuffler-0.6.0-10.fc26.noarch
3/3
Vérification de : pdfshuffler-0.6.0-10.fc26.noarch
1/3
Vérification de : pyPdf-1.13-13.fc26.noarch
2/3
Vérification de : pypoppler-0.12.2-5.fc26.x86_64
3/3

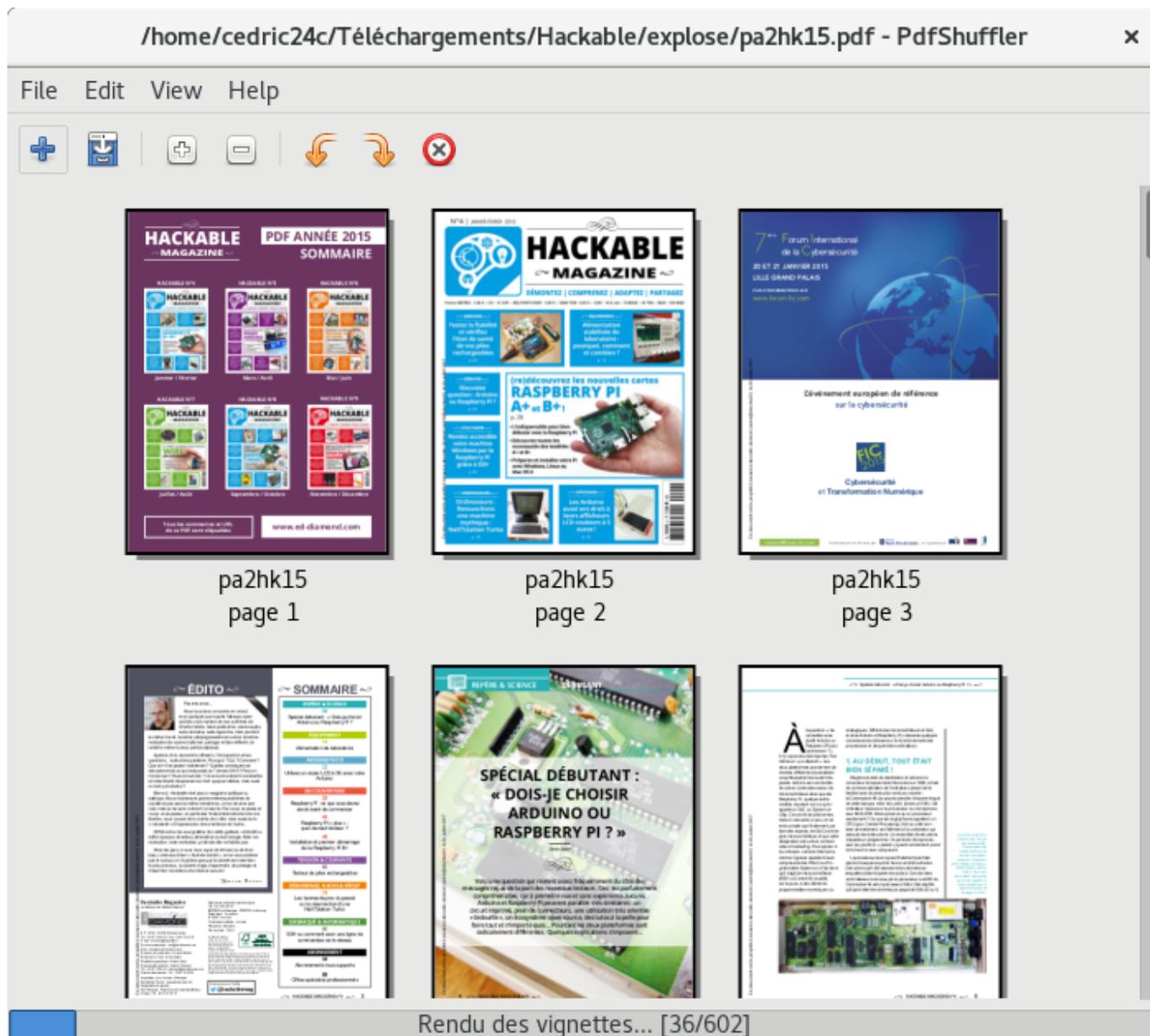
Installé :

pdfshuffler.noarch 0.6.0-10.fc26
pypoppler.x86_64 0.12.2-5.fc26

pyPdf.noarch 1.13-13.fc26

Terminé !

Utiliser pdf shuffler



Poppler

Installer poppler



Exécuter l'installation :

```
sudo dnf install poppler-utils
```

Utiliser poppler

poppler-utils est une collection d'outils qui s'appuie sur l'API **poppler** pour éditer et modifier les fichiers PDF.

- **pdfdetach** extract embedded documents from a PDF
- **pdffonts** lists the fonts used in a PDF
- **pdfimages** extract all embedded images at native resolution from a PDF
- **pdfinfo** list all infos of a PDF
- **pdfseparate** extract single pages from a PDF
- **pdftocairo** convert single pages from a PDF to vector or bitmap formats using cairo
- **pdftohtml** convert PDF to HTML format retaining formatting
- **pdftoppm** convert a PDF page to a bitmap
- **pdftops** convert PDF to printable PS format
- **pdftotext** extract all text from PDF
- **pdfunite** merges several PDF

The poppler-utils package provides several different commands for interacting with and modifying PDF files. However, the two commands that help us the most here are **pdfseparate** and **pdfunite**.

pdfseparate extracts pages into multiple PDFs that we can later merge together with **pdfunite**. To extract all the pages of a document into individual files, use:

```
pdfseparate ColoringBook.pdf ColoringBook-page_%d.pdf
```

To export a range of pages — say just pages 3,4,5,6,7,8 and 9 of a PDF — use the command:

```
pdfseparate -f 3 -l 9 ColoringBook.pdf ColoringBook-page_%d.pdf
```

Finally, after using **pdfseparate**, if we wanted to create a new document (NewColoringBook.pdf) with ColoringBook-page_3.pdf and ColoringBook-page_7.pdf we could use the following **pdfunite** command:

```
pdfunite ColoringBook-page_3.pdf ColoringBook-page_7.pdf NewColoringBook.pdf
```

PDFTK

Installer le programme PDFTK

```
yum install pdftk
```

Utiliser PDFTK

Assembler (merge, join) plusieurs fichiers PDF en un seul nouveau fichier PDF Syntaxe

```
pdftk P1 cat output P2
```

- P1 - source - correspond à la liste de fichiers à assembler
- P2 - destination - correspond au fichier créé avec les documents PDF assemblés.

Exemple

Assembler in1.pdf et in2.pdf dans un nouveau fichier out1.pdf :

```
pdftk in1.pdf in2.pdf cat output out2.pdf
```

Remarque

Si PDFMERGE est installé sur votre poste, il remplace aussi cette fonctionnalité :

```
pdfmerge file1.pdf file2.pdf... fileN.pdf outfile.pdf
```

PDF Chain

PDF Chain est une interface graphique pour PDF Toolkit (pdftk). Cette interface supporte toutes les commandes disponibles dans PDF Toolkit.

ImageMagick

Installer ImageMagick

Fedora 30

Sous **Fedora 30**, il suffit d'exécuter **dnf** pour installer le paquet présent dans des dépôts **fedora** et **updates** .

```
$ sudo dnf install ImageMagick
```

Les dépendances à d'autres paquets sont les suivantes :

```
provider: ImageMagick-libs-1:6.9.10.75-1.fc30.x86_64  
provider: glibc-2.29-27.fc30.x86_64  
provider: glibc-2.29-27.fc30.i686
```

1. La compil' des questions

1.1 Recompiler un fichier PDF (pdftocairo et poppler)

Dans un dossier vide, copier le fichier PDF à traiter. Se placer dans ce dossier avec la commande `cd` puis exécuter les commandes suivantes.

Convertir le fichier PDF en plusieurs fichiers JPEG avec **poppler** :

```
$ pdftocairo `ls -v *.jpg` -jpeg
```

Convertir les fichiers JPEG et les assembler en un fichier PDF avec **ImageMagick** :

```
$ convert `ls -v *.jpg` "final.pdf"
```

Effacer les fichiers temporaires :

```
$ rm *.jpg
```

Le fichier `final.pdf` est prêt.

1.2 Recompiler un fichier PDF (gs via ps)

Voici une méthode avec les programmes **pdf2ps** et **ps2pdf** en ligne de commande. Ces programmes sont intégrés au paquets **GhostScript**.

On considère que l'on souhaite transformer le fichier nommé **livre.pdf**. Le résultat sera dans le fichier **livre-opti.pdf**

```
pdf2ps livre.pdf inter.ps  
ps2pdf -dPDFSETTINGS=/prepress inter.ps livre-opti.pdf
```

L'option **dPDFSETTINGS** définit la qualité d'image issue du fichier PDF. Elle peut prendre les valeurs (du moins bon au meilleur résultat) **/screen**, **/default**, **/ebook**, **/printer**, **/prepress**

1.3 Recompiler un fichier PDF (gs)

Voici une méthode avec les programmes **gs** en ligne de commande. Le programme **gs** s'appelle **GhostScript** qui est un interpréteur de langage PDF et PS.

On considère que l'on souhaite transformer le fichier nommé **livre.pdf**. Le résultat sera dans le fichier **livre-opti.pdf**

```
gs -dBATCH -dNOPAUSE -q -sDEVICE=pdfwrite -dPDFSETTINGS=/ebook -sOutputFile=livre-opti.pdf livre.pdf
```

L'option **dPDFSETTINGS** définit la qualité d'image issue du fichier PDF. Elle peut prendre les valeurs (du moins bon au meilleur résultat) **/screen**, **/default**, **/ebook**, **/printer**, **/prepress**

1.4 Extraire le texte présent dans un PDF

Votre fichier PDF peut contenir un ensemble de textes. Celui-ci peut être extrait directement en fichier texte. Voici une méthode avec le programme **pdftotext** en ligne de commande. Ce programme fait partie de la suite **poppler-utils**.

On considère que l'on souhaite extraire le texte présent dans le fichier nommé **livre.pdf**. Le résultat sera dans le fichier **livre.txt**

```
pdftotext livre.pdf livre.txt
```

1.5 Deviner le texte d'un fichier PDF

Si votre fichier PDF ne comporte que des images, il est nécessaire de deviner le texte présent. Il s'agit de la méthode OCR. Voici une méthode avec le programme **ocrmypdf** en ligne de commande.

Ce programme peut être installé avec la commande :

```
sudo apt install ocrmypdf
```

On considère que l'on souhaite extraire le texte présent dans le fichier nommé **livre.pdf**. Le résultat sera dans le fichier **livre.txt**

```
ocrmypdf livre.pdf livre.txt -l fra
```

Le dictionnaire OCR s'appuie sur le programme **tesseract**. Il faut installer le dictionnaire que l'on souhaite utiliser par la suite :

```
sudo apt install tesseract-ocr-fra
```

From:
<https://www.abonnel.fr/> - **notes informatique & technologie**

Permanent link:
https://www.abonnel.fr/informatique/divers/manipuler_les_pdf_sous_linux

Last update: **2023/02/28 21:02**

