

pdftools

Urs Roesch

Version 0.6

Table of Contents

img2pdf.....	1
Usage.....	1
Examples	1
ocrpdf	2
Usage.....	2
Examples	2
pdflatex	4
Usage.....	4
Examples	4
pdfmeta.....	5
Usage.....	5
Examples	5
pdfsize.....	6
Usage.....	6
Examples	6

A collection of PDF command line tools and wrappers for Linux

img2pdf

A script to convert PNGs, TIFFs or JPEGs to PDF files.

License

MIT

Requires

bash, pdfcats, ImageMagick, pdftk

Usage

```
img2pdf <options> <img-file> [<img-file> ... ]
```

Options:

-h		--help	This message
-d		--delete	Delete the images after creating the PDF file.
-o		--output <name>	Write the output to specified file <name>.
			Implies -s
-r		--rotate <value>	Rotate the image by <value>
			Where value can be a positive or negative integer between 0 and 360.

Examples

Create single page pdfs from images

```
img2pdf first.png second.jpg
```

Create single page pdfs from images and then delete the original images.

```
img2pdf --delete first.png second.jpg
```

Rotate the image 180 degrees prior to creating the pdf

```
img2pdf --rotate 180 myimage.png
```

*Rotate the images by 180 degrees and concatenate everything into file **output.pdf***

```
img2pdf --output output.pdf --rotate 180 page1.png page2.jpg
```

ocrpdf

Runs PDFs through OCR and saves the output as a text searchable PDF with the same name.



Only works with quality degradation for PDFs comprised of a single JPEG, LZW or deflated image per page. LZW compressed images will be converted to deflated ones during the OCR process but there is no loss of fidelity. Composite images with multiple layers as JBIG2, CCITT or SMASKS are rendered as a flat image before running through OCR. The resulting PDFs are suffixed with '-ocrpdf.pdf' preserving the original file.

License

MIT

Requires

bash, pdfcats, pdftk, tesseract, pdftimages (poppler-utils), pdftocairo (poppler-utils)

Usage

```
ocrpdf [options] <file> [<file> [, ,]]
```

Options:

-R --render 	Render composite images as either png or jpeg before running through OCR. default: png
-h --help	This message
-v --version	Print version information and exit
-l --lang	Set the OCR language for multiple concatenate with a + E.g eng+deu for English and German default: osd+ita+deu+equ+fra+eng+jpn

Examples

Run OCR with all installed languages on a couple of PDFs

```
ocrpdf first.pdf second.pdf
```

Run OCR with German dictionary on a single PDF

```
ocrpdf --lang deu german.pdf
```

Run OCR with German, French and English dictionaries on multiple PDFs

```
ocrpdf --lang deu+fra+eng scanned_*.pdf
```

Convert composite images to JPEG before running running through OCR

```
ocrpdf --render jpeg scanned_document.pdf
```

pdfcat

A quick hack to replace **pdfunite** as it destroys too much of the original's meta data.

License

MIT

Requires

bash, pdftk >= 2.0

Usage

```
pdfcat <pdf> <pdf> [...]
```

Examples

Merging two PDFs into a new one

```
pdfcat first.pdf second.pdf > merged.pdf
```

Merging sequentially ordered pdfs into a single document

```
pdfcat myscan*.pdf > merged.pdf
```

pdfmeta

A wrapper script around `pdftk` to manipulate a PDFs meta data

License

MIT

Requires

bash >= 4.0, pdftk >= 2.0

Usage

```
pdfmeta <options> <pdf> [[<pdf>] ..]
```

Options:

-h --help	This message
-k --keywords	Comma seperated list of keywords
-s --subject	Define the PDFs subject
-t --title	Define the PDFs title
-c --creator	Define the PDFs creator program or library
-p --producer	Define the PDFs producing program
-C --creation-date	Set the creation date of the PDF
-M --modification-date	Set the modification date of the PDF

Examples

Modify keywords

```
pdfmeta --keywords "rainbow, magical, unicorn" unicorn.pdf rainbow.pdf
```

Modify creation date

```
pdfmeta --creation-date "2017-01-01 22:30:45" unicorn.pdf
```

pdfresize

A wrapper around `ghostscript` to reduce the size of a scanned document



pdfresize is very likely not working with PDF documents containing [JBIG2](#) images

License

MIT

Requires

bash, ghostscript

Usage

Usage: pdfresize [pdfsettings] <input> <output>

<input> A PDF file preferably of high resolution

<output> Name of the PDF file to save the result to

pdfsettings

screen - low-resolution; comparable to "Screen Optimized" in Acrobat Distiller

ebook - medium-resolution; comparable to "eBook" in Acrobat Distiller

printer - comparable to "Print Optimized" in Acrobat Distiller

prepress - comparable to "Prepress Optimized" in Acrobat Distiller

default - intended to be useful across a wide variety of uses

Examples

Resize to default resolution

```
pdresize input.pdf output.pdf
```

Resize to screen resolution

```
pdfresize screen input.pdf output.pdf
```