pdfcropper

Release 1.0.0

Sep 27, 2017

Contents

1 Česká verze 3

2 Script 5

3 API 7

4 Source code 11

5 Installation 13

6 Testing 15

Python Module Index 17

This project is aimed to provide simple and easy to use script for cropping PDF files as part of the E-deposit project.

CHAPTER 1

Česká verze

- /index_cz

CHAPTER 2

Script

From user's point of view, following script is his interface with pdfcropper:

pdf_cropper.py

Help:

```
usage: pdf_cropper.py [-h] [-c X X X X] [-e X X X X] [-r PAGE [PAGE ...]]
                      [-o OUTPUT]
                      filename

PDF cropper. This program can be used to crop and remove pages from PDF files.

positional arguments:
  filename              Input filename.

optional arguments:
  -h, --help            show this help message and exit
  -c X X X X, --crop X X X X
                        Crop vector for all or even pages. Vector is in format
                        LEFT RIGHT TOP BOTTOM. -c 50 50 10 10 for example.
  -e X X X X, --crop-odd X X X X
                        Crop vector for odd pages. Vector is in format LEFT
                        RIGHT TOP BOTTOM. -c 50 50 10 10 for example.
  -r PAGE [PAGE ...], --remove PAGE [PAGE ...]
                        Remove following pages. Page numbers starts from zero.
  -o OUTPUT, --output OUTPUT
                        Save modified file to given destination. If not set,
                        suffix '_cropped.pdf' is used.
```

Example of the use:

```
./pdf_cropper.py -c 10 10 5 5 -r 0 -- ukazka01b.pdf
```

Which will crop the file `ukazka01b.pdf` ten millimeters from left and right and five millimeters from up and down. `-r` or `--remove` parameter will also remove page at index `0`. Result can be seen in `ukazka01b_cropped.pdf`.

Česká verze

- /scripts/pdf_cropper_cz

CHAPTER 3

API

From programmer's point of view, there are few collections of functions, which he (or she) may use:

pdfcropper package

Submodules

cropper submodule

This module contains wrappers over pyPdf, which provides high level API for cropping PDF files.

Note: All sizes used in this module should be in millimeters.

API

`pdfcropper.cropper.`**`POINT_MM`** = **0.35277777777777775**

1pt = inch/72, 1 inch = 25.4 mm

`pdfcropper.cropper.`**`get_width_height`**(*page*)

Return width and height of the *page*.

Parameters **`page`** (*obj*) – `PdfFileReader.pages` instance.

Returns `(width, height)` as float, in millimeters.

Return type tuple

`pdfcropper.cropper.`**`crop_page`**(*page*, *left*, *right*, *top*, *bottom*)

Crop *page* to size given by *left*, *right*, *top* and *bottom*.

Parameters

- **page** (*obj*) – `pyPdf` PdfFileReader's page object.
- **left** (*int*) – Cut X millimeters from left.
- **right** (*int*) – Cut X millimeters from right.
- **top** (*int*) – Cut X millimeters from top.
- **bottom** (*int*) – Cut X millimeters from bottom.

> **Warning:** This functions modifies the *page* reference!

Returns Modified page object.

Return type obj

`pdfcropper.cropper.`**`crop_all`**(*pdf*, *left*, *right*, *top*, *bottom*, *remove=[]*)

Crop all pages in *pdf*. Remove pages specified by *remove*.

Parameters

- **pdf** (*obj*) – `pyPdf` `PdfFileReader` object.
- **left** (*int*) – Cut X millimeters from left.
- **right** (*int*) – Cut X millimeters from right.
- **top** (*int*) – Cut X millimeters from top.
- **bottom** (*int*) – Cut X millimeters from bottom.
- **remove** (*list/tuple, default []*) – List of integers. As the function iterates thru the pages in *pdf*, indexes of the pages which matchs those in *remove* will be skipped.

Returns `PdfFileWriter` instance, with modified pages.

Return type obj

`pdfcropper.cropper.`**`crop_differently`**(*pdf*, *even_vector*, *odd_vector*, *remove=[]*)

Crop *pdf* even pages by *even_vector* and odd pages by *odd_vector*. Remove pages specified by *remove*.

Parameters

- **pdf** (*obj*) – `pyPdf` `PdfFileReader` object.
- **even_vector** (*list*) – List of coordinates to which all even pages will be cropped. `[Left, Right, Top, Bottom]`.
- **odd_vector** (*list*) – List of coordinates to which all odd pages will be cropped. `[Left, Right, Top, Bottom]`.
- **remove** (*list/tuple, default []*) – List of integers. As the function iterates thru the pages in *pdf*, indexes of the pages which matchs those in *remove* will be skipped.

Returns `PdfFileWriter` instance, with modified pages.

Return type obj

`pdfcropper.cropper.`**`remove_pages`**(*pdf*, *remove*)

Remove pages specified in vector *remove*.

Parameters

- **pdf** (*obj*) – `pyPdf` `PdfFileReader` object.

- **remove** (*list/tuple, default []*) – List of integers. As the function iterates thru the pages in *pdf*, indexes of the pages which matchs those in *remove* will be skipped.

Returns `PdfFileWriter` instance, with modified pages.

Return type obj

`pdfcropper.cropper.`**`read_pdf`**(*filename*)

Read pdf file specified by *filename*.

Parameters **filename** (*str*) – Path to the pdf file.

Returns `PdfFileReader` object.

Return type obj

`pdfcropper.cropper.`**`save_pdf`**(*filename*, *content*)

Save *content* to *filename*.

Parameters

- **filename** (*str*) – Path which will be used for *content*.
- **content** (*obj*) – `PdfFileWriter` object which will be serialized.

CHAPTER 4

Source code

This project is released as opensource (GPL) and source codes can be found at GitHub:

- https://github.com/edeposit/pdfcropper

CHAPTER 5

Installation

Module is hosted at PYPI, and can be easily installed using PIP:

```
sudo pip install pdfcropper
```

CHAPTER 6

Testing

Almost every feature of the project is tested in unit/integration tests. You can run this tests using provided `run_tests.sh` script, which can be found in the root of the project.

Requirements

This script expects that pytest is installed. In case you don't have it yet, it can be easily installed using following command:

```
pip install --user pytest
```

or for all users:

```
sudo pip install pytest
```

Python Module Index

p

pdfcropper, 7
pdfcropper.cropper, 7

Index

C

crop_all() (in module pdfcropper.cropper), 8
crop_differently() (in module pdfcropper.cropper), 8
crop_page() (in module pdfcropper.cropper), 7

G

get_width_height() (in module pdfcropper.cropper), 7

P

pdfcropper (module), 7
pdfcropper.cropper (module), 7
POINT_MM (in module pdfcropper.cropper), 7

R

read_pdf() (in module pdfcropper.cropper), 9
remove_pages() (in module pdfcropper.cropper), 8

S

save_pdf() (in module pdfcropper.cropper), 9