Red Stork

Release 0.0.1

Apr 02, 2020

Manual:

Yet another PDF parser. This one is based on PDFium engine.

Quick Start

Sample:

```python
from redstork.document import Document

doc = Document('sample.pdf')
print('Number of pages:', len(doc))
```

1.1 Tutorial

TODO

1.2 API Reference

1.2.1 Document

class redstork.**Document**(*file_name*, *password=None*)
    PDF document.

    A list-like container of pages. Sample use:

```python
doc = Document('sample.pdf')
print("Number of pages:', len(doc))

for key, value in doc.meta.items():
    print('    ', key, ':', value)
```

__init__(*file_name*, *password=None*)
    Create new PDF Document object, from a file.

        Parameters

> - **file_name** (*str*) – Name of PDF file
>
> - **password** (*str*) – File password (optional)

numpages = None
> *int* – total number of pages

meta = None
> *dict* – document meta info (Author, Title, etc)

__getitem__(*page_index*)
> Returns page at this index.
>
> Example:

```
doc = ...

page = doc[0]  # first page
```

> > **Parameters page_index** (*int*) – zero-based page index
> >
> > **Returns** *Page* object

__len__()
> Returns number of pages in this document

1.2.2 Page

class redstork.**Page**(*page*, *page_index*, *parent*)
> Represents page of a PDF file.

crop_box
> Page crop box.

media_box
> Page media box.

rotation
> Page rotation.
>
> - 0 - no rotation
>
> - 1 - rotated 90 degrees clock-wise
>
> - 2 - rotated 180 degrees clock-wise
>
> - 3 - rotated 270 degrees clock-wise

label
> Page label.

__len__()
> Number of objects on this page.

__getitem__(*index*)
> Get object at this index.

__iter__()
> Iterates over page objects.

render(*file_name*, *scale=1.0*, *rect=None*)
> Render page (or rectangle on the page) as PPM image file.

> **Parameters**
>
> - **file_name** (`str`) – name of the output file
>
> - **scale** (`float`) – scale to use (default is 1.0, which will assume that 1pt takes 1px)
>
> - **rect** (`tuple`) – optional rectangle to render. Value is a 4-tuple of (x0, y0, x1, y1) in PDF coordinates. if None, then page's `crop_box` will be used for rendering.

1.2.3 Font

class redstork.**Font** (*font*, *parent*)
    Represents font used in a PDF file.

    FLAGS_NORMAL = 0
        Normal font

    FLAGS_FIXED_PITCH = 1
        Fixed pitch font

    FLAGS_SERIF = 2
        Serif font

    FLAGS_SYMBOLIC = 4
        Symbolic font

    FLAGS_SCRIPT = 8
        Script font

    FLAGS_NONSYMBOLIC = 32
        Non-symbolic font

    FLAGS_ITALIC = 64
        Italic font

    FLAGS_ALLCAP = 65536
        All-cap font

    FLAGS_SMALLCAP = 131072
        Small-cap font

    FLAGS_FORCE_BOLD = 262144
        Force-bold font

    name
        Font name in the PDF document.

    flags
        Font flags.

    weight
        Font weight.

CHAPTER 2

Indices and tables

- genindex
- modindex
- search
- Glossary

Python Module Index

r

redstork, 3

Index

Symbols