Package Index > **pdfminer3k** > **1.3.1**

pdfminer3k 1.3.1

PDF parser and analyzer

Download
pdfminer3k-1.3.1.tar.gz

pdfminer3k is a Python 3 port of pdfminer. PDFMiner is a tool for extracting information from PDF documents. Unlike other PDF-related tools, it focuses entirely on getting and analyzing text data. PDFMiner allows to obtain the exact location of texts in a page, as well as other information such as fonts or lines. It includes a PDF converter that can transform PDF files into other text formats (such as HTML). It has an extensible PDF parser that can be used for other purposes instead of text analysis.

Changes

Version 1.3.1 – 2016/11/05

- Replaced root loggers with module-wide loggers. This allows user to disable the log messages from pdfminer3k.

Version 1.3.0 – 2012/07/20

- Added *pdfexplore*, a tool to debug PDFs by exploring their data.
- Don't try to group textboxes when there's too many (it takes too long).
- Support object references as filters in streams.
- Parse every object as soon as an objectid can't be found.
- Improved the *STRICT*-based error handling idiom.

Version 1.2.4 – 2011/10/07

- When xref tables are corrupt, parse and cache all objects as a fallback.
- Fixed a bogus assertion in layouts.

Version 1.2.3 – 2011/09/05

- Fixed a crash on uneven cmap codes.
- Fixed a meta-crash caused by bad PSParser repr.

Version 1.2.2 – 2011/08/30

- Fixed crash on corrupt LZW data.
- Ignore lines with no text for textlines grouping.
- Don't crash on invalid dictionary constructs when parsing postscript.

Version 1.2.1 – 2011/08/22

- Fixed a crash on corrupted inline images.
- Tweaked layout detection algo.

Version 1.2.0 – 2011/08/09

- There wasn't a changelog until now. Starting it.
- Removed the old Postscript lexer and replaced it by a PLY-based one.
- Added a couple of heuristic layout features.
- Fixed a couple of crashes on opening PDFs.

File	Type	Py Version	Uploaded on	Size
pdfminer3k-1.3.1.tar.gz (md5)	Source		2016-10-05	3MB

Author: Jaepil Jeong, Virgil Dupras
Home Page: https://github.com/jaepil/pdfminer3k
Keywords: pdf parser
License: MIT/X
Categories
    Development Status :: 4 - Beta
    Environment :: Console
    Intended Audience :: Developers
    Intended Audience :: Science/Research
    License :: OSI Approved :: MIT License
    Topic :: Text Processing
Package Index Owner: hsoft, jaepil
DOAP record: pdfminer3k-1.3.1.xml