|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 5/17/24, 1:56 PM | [Sponsors](https://pypi.org/sponsors/) | Search projects | | pdf2docx · PyPI |  |
|  |
| [Help](https://pypi.org/help/) | [Log in](https://pypi.org/account/login/) | [Register](https://pypi.org/account/register/) |

pdf2docx 0.5.8

[Latest version](https://pypi.org/project/pdf2docx/)

|  |  |  |
| --- | --- | --- |
| pip install pdf2docx |  | Released: Jan 22, 2024 |

Open source Python library converting pdf to docx.

|  |  |
| --- | --- |
| Navigation   Project  description   Release | Project description  English | [中文](https://pypi.org/project/pdf2docx/README_CN.md)  pdf2docx |

history

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | python >=3.6 | |  | | --- | |  | | [codecov](https://codecov.io/gh/dothinking/pdf2docx) | |  | | --- | | 87% | | [pypi](https://pypi.python.org/pypi/pdf2docx/) [v0.5.8](https://pypi.python.org/pypi/pdf2docx/) | license GPL v3 |
|  Download | downloads 152k/month | | |  |  |  |

files

|  |  |  |
| --- | --- | --- |
|  | Extract data from PDF with PyMuPDF , e.g. text, images and   |  | | --- | |  | |

drawings

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | | | | |
| Verified details | |  | Parse layout with rule, e.g. sections, paragraphs, images and | |  |
| These details have been | | tables | | |
| verified by PyPI | |  | Generate docx with python-docx   |  | | --- | |  | | |
| Maintainers | | Features | | |
|  | [dothinking](https://pypi.org/user/dothinking/) |  | Parse and re-create page layout | |
|  | | page margin |
| Unverified details | |  | | section and column (1 or 2 columns only) | 1/4 |
|  | | page header and footer [TODO] |
| https://pypi.org/project/pdf2docx/ | |



|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 5/17/24, 1:56 PM | | | | |  | pdf2docx · PyPI | |
| These details have not been | | | | | Parse and re-create paragraph | |
| verified by PyPI | | | | |  |  | OCR text [TODO] |
|  | text in horizontal/vertical direction: from left to right, from |
| Project links | | | | |
| bottom to top | |
|  | [Homepage](https://artifex.com/) | | | |
|  | font style, e.g. font name, size, weight, italic and color |
| View statistics for this | | | | |  | text format, e.g. highlight, underline, strike-through |
| project via | | | | |
|  | list style [TODO] |
| [Libraries.io](https://libraries.io/pypi/pdf2docx) | | | [](https://libraries.io/pypi/pdf2docx), or by | |
|  | external hyper link |
| using [our public](https://packaging.python.org/guides/analyzing-pypi-package-downloads/) | | | | |
|  | paragraph horizontal alignment (left/right/center/justify) |
| [dataset on Google](https://packaging.python.org/guides/analyzing-pypi-package-downloads/) | | | | |
| [BigQuery](https://packaging.python.org/guides/analyzing-pypi-package-downloads/) | | [](https://packaging.python.org/guides/analyzing-pypi-package-downloads/) | | | and vertical spacing | |
| Parse and re-create image | |
| Meta | | | | |
| License: GPL v3 | | | | |  |  | in-line image |
|  | image in Gray/RGB/CMYK mode |
| Author: [Artifex](mailto:support@artifex.com) | | | | [](mailto:support@artifex.com) |
|  | transparent image |
|  | floating image, i.e. picture behind text |
| pdf-to-word, pdf- | | | | |
| Parse and re-create table | |
| to-docx | | | | |
| Requires: Python | | | | |  | border style, e.g. width, color |
|  | shading style, i.e. background color |
| >=3.6 | | | | |
|  | merged cells |
| [](https://server.ethicalads.io/proxy/click/3348/208a9923-9d80-4936-ba23-8756cf3cd9bd/) | | | | |  |  | vertical direction cell |
|  |  | table with partly hidden borders |
| Indeed is a | | | | |  | nested tables |
| Contributing | | | | | Parsing pages with multi-processing | |

sponsor of the

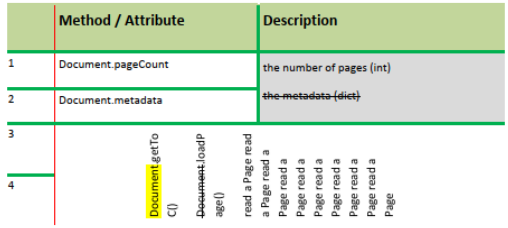
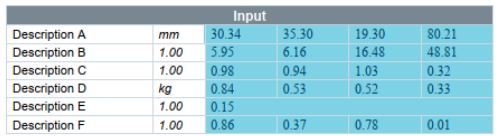
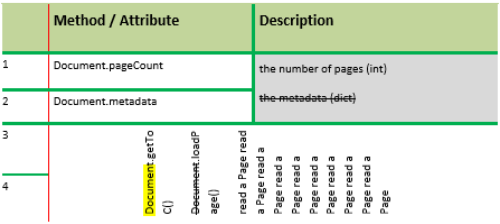
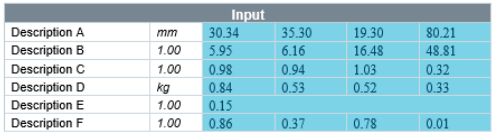
|  |  |
| --- | --- |
| Python Software Foundation. | It can also be used as a tool to extract table contents since both table content and format/style is parsed. |

[PSF Sponsor](https://www.python.org/psf/sponsorship/?ref=ethicalads-placement) · [Served](https://www.ethicalads.io/sponsorship-platform/?ref=psf)

|  |  |  |
| --- | --- | --- |
| [ethically](https://www.ethicalads.io/sponsorship-platform/?ref=psf) | Limitations | |
|  |  | Text-based PDF file |
|  | Left to right language |
|  | Normal reading direction, no word transformation / rotation |
|  | Rule-based method can't 100% convert the PDF layout |

Documentation

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| https://pypi.org/project/pdf2docx/ |  | [Installation](https://pdf2docx.readthedocs.io/en/latest/installation.html) | | 2/4 |
|  | [Quickstart](https://pdf2docx.readthedocs.io/en/latest/quickstart.html) | |
|  | [Convert PDF](https://pdf2docx.readthedocs.io/en/latest/quickstart.convert.html) |



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 5/17/24, 1:56 PM |  |  | [Extract table](https://pdf2docx.readthedocs.io/en/latest/quickstart.table.html) | pdf2docx · PyPI |
|  | [Command Line Interface](https://pdf2docx.readthedocs.io/en/latest/quickstart.cli.html) |
|  | [Graphic User Interface](https://pdf2docx.readthedocs.io/en/latest/quickstart.gui.html) |
| [Technical Documentation (In Chinese)](https://pdf2docx.readthedocs.io/en/latest/techdoc.html) | |
|  | [API Documentation](https://pdf2docx.readthedocs.io/en/latest/modules.html) | |

Sample





|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Help | | | | | About PyPI | | [](https://dtdg.co/pypi) |
| [Installing packages](https://packaging.python.org/tutorials/installing-packages/) | | [](https://packaging.python.org/tutorials/installing-packages/) | | | [PyPI Blog](https://blog.pypi.org/) | [](https://blog.pypi.org/) |
| [Uploading packages](https://packaging.python.org/tutorials/packaging-projects/) | | | [](https://packaging.python.org/tutorials/packaging-projects/) | | [Infrastructure dashboard](https://dtdg.co/pypi) | |
| [Statistics](https://pypi.org/stats/) | |
| [User guide](https://packaging.python.org/) | [](https://packaging.python.org/) | | | |
| [Project name retention](https://www.python.org/dev/peps/pep-0541/) | | | | [](https://www.python.org/dev/peps/pep-0541/) | [Logos & trademarks](https://pypi.org/trademarks/) | |
| [FAQs](https://pypi.org/help/) | | | | | [Our sponsors](https://pypi.org/sponsors/) | |
| Contributing to PyPI | | | | | Using PyPI | |

[Bugs and feedback](https://pypi.org/help/#feedback)   
[Contribute on GitHub](https://github.com/pypi/warehouse)

[Code of conduct](https://github.com/pypa/.github/blob/main/CODE_OF_CONDUCT.md)

[Report security issue](https://pypi.org/security/)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| https://pypi.org/project/pdf2docx/ | Translate PyPI | [](https://hosted.weblate.org/projects/pypa/warehouse/) | Privacy policy | [](https://www.python.org/privacy/) | 3/4 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 5/17/24, 1:56 PM | pdf2docx · PyPI | | | |
| [Sponsor PyPI](https://pypi.org/sponsors/) | [Terms of use](https://pypi.org/policy/terms-of-use/) | | |
| [Development credits](https://github.com/pypi/warehouse/graphs/contributors) | [](https://github.com/pypi/warehouse/graphs/contributors) | [Acceptable Use Policy](https://pypi.org/policy/acceptable-use-policy/) | |
| Status: Service Under Maintenance | | | [](https://status.python.org/) |

Developed and maintained by the Python community, for the Python community. [Donate today!](https://donate.pypi.org/)

"PyPI", "Python Package Index", and the blocks logos are registered [trademarks](https://pypi.org/trademarks/) of the [Python Software Foundation](https://www.python.org/psf-landing).

© 2024 [Python Software Foundation](https://www.python.org/psf-landing/)

[Site map](https://pypi.org/sitemap/)

Switch to desktop version

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | English | español | français | 日本語 | português (Brasil) | українська | Ελληνικά | Deutsch | 中文 (简体) |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 中文 (繁體) | русскийעבריתEsperanto | |  |
|  |  |  |  |  |
| AWS | Datadog | Fastly | Google | Microsoft |
| Cloud computing |

[and Security](https://aws.amazon.com/)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sponsor | Monitoring | CDN | Download Analytics | PSF Sponsor | 4/4 |
|  |
|  |  |
| [StatusPage](https://statuspage.io/) |
| [Pingdom](https://www.pingdom.com/) | [Sentry](https://getsentry.com/for/python) |
| [Monitoring](https://www.pingdom.com/) | [Error logging](https://getsentry.com/for/python) | [Status page](https://statuspage.io/) |
| https://pypi.org/project/pdf2docx/ |