

Convert PDF

We can use either the `Converter` class, or a wrapped method `parse()` to convert all/specified pdf pages to docx. Multi-processing is supported in case pdf file with a large number of pages.

Example 1: convert all pages

```
from pdf2docx import Converter

pdf_file = '/path/to/sample.pdf'
docx_file = 'path/to/sample.docx'

# convert pdf to docx
cv = Converter(pdf_file)
cv.convert(docx_file)      # all pages by default
cv.close()
```

An alternative using `parse` method:

```
from pdf2docx import parse

pdf_file = '/path/to/sample.pdf'
docx_file = 'path/to/sample.docx'

# convert pdf to docx
parse(pdf_file, docx_file)
```

Example 2: convert specified pages

- Specify pages range by `start` (from the first page if omitted) and `end` (to the last page if omitted):

```
• # convert from the second page to the end (by default)
• cv.convert(docx_file, start=1)
•
• # convert from the first page (by default) to the third (end=3, excluded)
• cv.convert(docx_file, end=3)
•
• # convert from the second page and the third
• cv.convert(docx_file, start=1, end=3)
```

- Alternatively, set separate pages by `pages`:

```
• # convert the first, third and 5th pages
• cv.convert(docx_file, pages=[0,2,4])
```

Note

Refer to `convert()` for detailed description on the input arguments.

Example 3: multi-Processing³

Turn on multi-processing with default count of CPU:

```
cv.convert(docx_file, multi_processing=True)
```

Specify the count of CPUs:

```
cv.convert(docx_file, multi_processing=True, cpu_count=4)
```

Note

Multi-processing works for continuous pages specified by `start` and `end` only.

Example 4: convert encrypted pdf³

Provide `password` to open and convert password protected pdf:

```
cv = Converter(pdf_file, password)
cv.convert(docx_file)
cv.close()
```

[Previous](#)[Next](#)

© Copyright 2021, dothinking.

Built with [Sphinx](#) using a [theme](#) provided by [Read the Docs](#).