

レポートを PDF にまとめる自動処理

著者：関勝寿 公開日：2023 年 12 月 23 日 ソース：<https://sekika.github.io/2023/12/23/report-pdf/>

たとえば Python で自動処理をしてまとめた結果などを、少しでも見栄えをととのえて PDF にするまでを自動処理するための方法をまとめる。

- ・処理結果をマークダウン .md として出力する
 - ・マークダウンを pandoc でワードファイル .docx に変換する
 - ・ワードファイルを python-docx で調整する
 - ・ワードファイルを docx2pdf で PDF に変換して最終出力とする
- このように、いくつかのツールを組み合わせる。

なお、ワードでは十分に見栄えの調整ができない場合には、LaTeX を経由するのが良いであろう。Pandoc は LaTeX を扱うこともできる。

1. マークダウンから pandoc で Word へ

まずは、処理結果をマークダウンとして出力する。テキストファイルで出力された結果をマークダウンに変換するプログラムを書いても良い。

次に、マークダウンをワードファイルに変換する。そのためには pandoc を使う。まずは pandoc をインストールし、

```
pandoc report.md -o report.docx --reference-doc=template.docx
```

とする。ここで、report.md はマークダウンのファイル名、report.docx は出力ファイル名、template.docx はテンプレートファイルで、用紙の余白などを設定しておく。マークダウンの最低限の書式、たとえばリストがワードの形式に変換される。ここで、ワードファイルが思い通りの見栄えにならないときにはマークダウンの出力を調整してやり直す。

2.Word を python-docx で調整

マークダウンでは、たとえば特定の行を中央寄せしたり右寄せしたり、といった調整は難しい（マークダウン上で指定しても、それを pandoc で変換すると、ワードのファイルでは必ずしもそれが反映されない）。そこで、Python の python-docx パッケージで調整をする。

以下は、コマンドライン引数からファイル名を受け取り、その Word ファイルの 1 行目を中央寄せに、2 行目を右寄せに変換して保存するプログラムである。

```
#!/usr/bin/env python3
import sys
from docx import Document
from docx.enum.text import WD_ALIGN_PARAGRAPH
# コマンドライン引数からファイル名を取得
if len(sys.argv) < 2:
    print(" ファイル名を引数として指定してください。")
    sys.exit(1)
file_path = sys.argv[1]
# Word 文書を開く
doc = Document(file_path)
# 最初の段落を中央寄せに設定
if len(doc.paragraphs) > 0:
    doc.paragraphs[0].alignment = WD_ALIGN_PARAGRAPH.CENTER
# 次の段落を右寄せに設定
if len(doc.paragraphs) > 1:
    doc.paragraphs[1].alignment = WD_ALIGN_PARAGRAPH.RIGHT
# 文書を保存
doc.save(file_path)
```

3.Word を pdf へ

Python の docx2pdf ライブラリで変換できる。

4.Makefile の記述

以下に、以上の作業をするための Makefile の例を示す。style.py はワードのスタイルを調整するプログラムであり、たとえば上記のようなものである。たとえば report.md が新規作成されるか更新されれば、make によって report.md → report.docx → report.pdf と変換される。

```
all: report.pdf
```

```
%.docx: %.md
    pandoc $< -o $@ --reference-doc=template.docx
    ./style.py $@
```

```
%.pdf: %.docx
    docx2pdf $<
```