

# 语雀文档的 PDF 生成工具

钒钛智能

2020.03

## Contents

<b>1 缘起</b>	<b>3</b>
<b>2 流程</b>	<b>3</b>
<b>3 使用</b>	<b>4</b>
3.1 环境需求 . . . . .	4
3.2 初始化 . . . . .	4
3.3 运行 . . . . .	4
3.4 调整 . . . . .	5

## 1 缘起

语雀是一个“专业的云端知识库”，很适合组织使用，沉淀知识，书写文档，语雀最开始主要以 Markdown 的在线编辑为主，后来逐渐形成了自己的一些特性，并不完全兼容 Markdown。

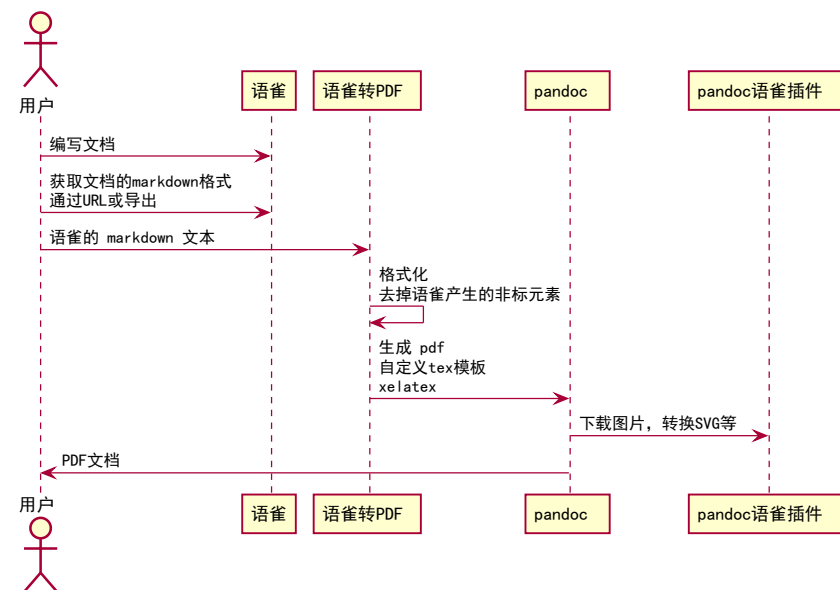
语雀前端技术主要为 H5 相关，这使得其产生的文档很适合浏览器阅读和分发，但在有时候，需要更加正式的可打印版本时，其虽然也提供了导出 Word、PDF 等功能，但基本还是 H5 格式的效果，缺少

- 页眉、页脚设置
- 目录
- 标题页

等“正式文档”的感觉。

对于 Markdown 文件，pandoc 有很好的工具链可以产生相应的 PDF，本工具就是利用这样的工具链来产生 PDF。

## 2 流程



在上面的流程中，本工具提供了如下的几类功能

- 格式化语雀产生的 Markdown，使其可以被 pandoc 使用

- 一个 pandoc 的[过滤器](#)，用于下载文档中的图片，以便插入到目标 PDF 中。特别的，对语雀文本绘图功能产生的 SVG 图像（plantuml、mermind、甘特图等），通过 svglib 将其转换成 PDF 后，使其在最终的 PDF 文件里，仍具备平滑缩放的效果。
- svglib 对中文的处理有一些问题，所以对其做了一些调整，使用 simhei 这个比较通用的字体来渲染 PDF。
- pandoc 默认的 tex 模板，对中文适配一般，所以做了一些调整，参看 fantai.tex 文件。

## 3 使用

### 3.1 环境需求

- Python3.6+ with pip
- 支持 xelatex 的 tex 发布版
  - windows: 在使用 scoop 的情况下，通过 scoop install latex 来安装
  - mac: 在使用 brew 的情况下，通过 brew install mactex 来安装
- pandoc 的较新版本
  - windows: 在使用 scoop 的情况下，通过 scoop install pandoc 来安装
  - mac: 在使用 brew 的情况下，通过 brew install pandoc 来安装

### 3.2 初始化

```
git clone https://github.com/elsejj/yueque2pdf.git
cd yueque2pdf
```

```
# 安装依赖的 python 库
pip install -r requirements.txt
```

### 3.3 运行

```
# 假设 tests/sample.md 是一个语雀产生的 Markdown 文件
```

```
python build.py tests/sample.md
```

```
# 如果一切顺利，那么会产生 tests/sample.pdf
```

### 3.4 调整

- 页眉 logo: logo.png 用于页眉 logo, 可以将其换为公司的 logo, 建议图像高度不超过 32px
- build.py
  - 替换 author 为需要的作者名
  - 替换 思源黑体 为需要的字体名
  - 替换 14pt 为需要的字体大小