

# README

南方小智

2020 年 8 月 10 日

# 目录

<b>第一章 概述</b>	<b>2</b>
1.1 处理单个文档 . . . . .	2
1.2 处理多个文档 . . . . .	3
1.3 生成图片文件 . . . . .	4
1.4 所有测试用例 . . . . .	4
<b>第二章 Markdown 支持</b>	<b>7</b>
2.1 特殊字符 . . . . .	7
2.2 导入代码文件 . . . . .	7
2.3 导入 tex 格式的图片文件 . . . . .	11
2.4 导入 dot 格式的图片文件 . . . . .	11
2.5 Markdown 的解析和渲染 . . . . .	12
<b>第三章 TODO</b>	<b>17</b>

# 第一章 概述

通过运行以下命令，可以将本 Markdown 文档生成 pdf 格式，建议阅读 pdf 版本的 README。

```
1 make readme
```

本工具用于个人辅助写作和作图等，满足各种文档编辑的需求。

- 该系统写作部分以基本的 Markdown 语法为核心，作图部分使用 Latex 和 dot 等通用工具包。
- 该系统会生成 Markdown 语法树，可通过 Latex 渲染生成 pdf 等，也可以通过其他渲染方式生成 Html, Word 等格式，方便发布文章。
- 该系统支持将批量的 Markdown 文档整理成册，比如合成一本书，并设置封面和章节目录等。
- 该系统可以通过 pgfplot, tikz 等工具，快速制作复杂的数学图片。

## 1.1 处理单个文档

本系统可以将一个 Markdown 文档转化为 tex 文档，并自动生成封面和目录，后续还通过 latex 工具包转化为 pdf。只需要将文档路径传入 path。

```
1 python3 create_book.py --path README.md --name README --author 南方小智
2 @echo 'xelatex cmd support Chinese'
3 xelatex -output-directory log README.tex
4 @echo 'run twice to build toc correctly'
5 xelatex -output-directory log README.tex
```

- name 代表生成的文件名和封面标题。
- author 代表生成的封面作者。

- 生成的 tex 文件在 log 文件夹中。

## 1.2 处理多个文档

本系统可以将多个 Markdown 文档合成一个 tex 文档，并自动生成封面和目录，后续还通过 latex 工具包转化为 pdf。只需要将所有文档放在同一个文件夹里，并将文件夹路径传入 path。

```
1 python3 create_book.py --path examples/趣题集/ --name 趣题集 --bg images/趣题集/background.jpeg --author 南方小智
2 @echo 'xelatex cmd support Chinese'
3 xelatex -output-directory log 趣题集.tex
4 @echo 'run twice to build toc correctly'
5 xelatex -output-directory log 趣题集.tex
```

- 支持多层文件夹，每层文件夹对应一个目录级别，比如文件夹内第一层目录每个文件夹名字为每章的名字，第二层目录每个文件夹名字为每节的名字。
- MAIN 文件为保留文件名，为该层目录对应章节的导言部分。
- \_INDEX 文件为保留文件名，用于对文件夹（章节）进行排序，目前章节的排列方法是较长的章节放在靠前的位置。
- 名字以“\_”开始的文件夹或者文件会被忽略。

\_INDEX 文件中按顺序列举该层目录的标题，未列举的标题将放在“\*”的位置。如果“\*”未标注在文件中，则未列举的标题默认排到最后。

```
1 概率统计
2 几何原本
3 逻辑思维
4 博弈游戏
5 数学皇后
6 魔法幻术
7 思想实验
8 思维锦囊
```

## 1.3 生成图片文件

本系统可以编译单一 tex 文档，并生成图片。只需要使用 simple 参数，就可以省略标题，目录，页码等，生成一个独立的图片形式，后续可以用 convert 命令将 pdf 转为图片。

```
1 python3 create_book.py --path images/趣题集/三角形悖论/image0.tex --name 三角
   形 --simple
2 xelatex -output-directory log 三角形.tex
3 @echo '需要安装brew install imagemagick'
4 convert -density 300 log/三角形.pdf -quality 90 log/三角形.png
```

## 1.4 所有测试用例

```
1 make readme      # 将本文档生成pdf
2 make all         # 生成一本《趣题集》
3 make image       # 生成一张png图片（需要安装brew install imagemagick）
4 make clean       # 清空log文件
5 make clean_all   # 清空所有生成文件，包括，log文件，tex文件， pdf文件等
```

具体细节请参照 makefile:

```
1 all:
2   python3 create_book.py --path examples/趣题集/ --name 趣题集 --bg images/趣
   题集/background.jpeg --author 南方小智
3   @echo 'xelatex cmd support Chinese'
4   xelatex -output-directory log 趣题集.tex
5   @echo 'run twice to build toc correctly'
6   xelatex -output-directory log 趣题集.tex
7   open log/趣题集.pdf
8
9 readme:
10  python3 create_book.py --path README.md --name README --author 南方小智
11  @echo 'xelatex cmd support Chinese'
12  xelatex -shell-escape -output-directory log README.tex
13  @echo 'run twice to build toc correctly'
14  xelatex -shell-escape -output-directory log README.tex
15  open log/README.pdf
```

```
16
17 image:
18   python3 create_book.py --path images/趣题集/三角形悖论/image0.tex --name 三
      角形 --simple
19   xelatex -output-directory log 三角形.tex
20   @echo '需要安装brew install imagemagick'
21   convert -density 300 log/三角形.pdf -quality 90 log/三角形.png
22   open log/三角形.png
23
24 test:
25   python3 test.py --path README.md
26
27 git:
28   git push https://github.com/JimmyFromSYSU/latex.git master
29
30 temp:
31   python3 create_book.py --path books/全栈开发项目实践/ --name 全栈开发项目实
      践 --author "" --output ignore
32   @echo 'xelatex cmd support Chinese'
33   xelatex -output-directory ignore 全栈开发项目实践.tex
34   @echo 'run twice to build toc correctly'
35   xelatex -output-directory ignore 全栈开发项目实践.tex
36   open ignore/全栈开发项目实践.pdf
37
38 clean:
39   rm -f log/*.aux
40   rm -f log/*.toc
41   rm -f log/*.log
42   rm -f log/*.out
43
44 clean_all:
45   rm -f log/*.aux
46   rm -f log/*.toc
47   rm -f log/*.log
48   rm -f log/*.out
49   rm -f log/*.tex
50   rm -f log/*.png
```

```
51  rm -f log/*.pdf
```

## 第二章 Markdown 支持

基本语法请参照：[markdown wiki](#)

### 2.1 特殊字符

在 Markdown 文件中使用特殊字符时需要进行特殊处理，比如添加转义字符 \

```
1 _ -> \_    # _ 在Markdown中可用于表示斜体
2 * -> \*    # * 在Markdown中可用于表示斜体
3 $ -> \$    # $ 在Markdown中可用于开始数学表达式模式
4 \ -> \\    # \ 是特殊的Latex符号
```

### 2.2 导入代码文件

```
1 {{create_book.py}}[code:Python] # 导入Python格式的代码文件

1 #!/usr/bin/python
2 # -*- coding: UTF-8 -*-
3 from src.render.latex import DOCUMENT, ABSTRACT, SECTION, NEW_LINE
4 from src.utils.path import read, write, trim_ext
5 from src.converter import md2tex
6 from src.constants import LOGGER_FORMAT
7 from os import listdir
8 from os.path import isfile, isdir
9 import markdown
10 import argparse
11 import logging
```



```

12 logger = logging.getLogger()
13 logging.basicConfig(level=logging.INFO, format=LOGGER_FORMAT)
14
15 MAIN_FILE = "MAIN"
16 INDEX_FILE = "_INDEX"
17
18
19 def get_title(section: str) -> str:
20     return trim_ext(section)
21
22
23 def create_by_folder(args) -> str:
24     def travel(path: str, level: int = 0) -> str:
25         # TODO: refactoring into a class
26         content = ""
27         sections = listdir(path)
28
29         # Add main content in this section before add sub sections.
30         if MAIN_FILE in sections:
31             child_path = path + "/" + MAIN_FILE
32             if isfile(child_path):
33                 file_content = md2tex(read(child_path), level + 1)
34                 content += NEW_LINE([file_content])
35
36         str_sections = []
37         sections_titles = []
38         for section in sections:
39             # ignore this section for now if it's prefix with '_'
40             if len(section) > 0 and section[0] in ['_', '.']:
41                 continue
42             child_path = path + "/" + section
43             if isfile(child_path) and section != MAIN_FILE:
44                 file_content = md2tex(read(child_path), level + 1)
45                 str_sections.append(SECTION(section, file_content, level))
46                 sections_titles.append(section)
47             elif isdir(child_path):
48                 str_section = NEW_LINE([

```

```

49         SECTION(section, "", level),
50         travel(child_path, level=level+1)
51     ])
52     str_sections.append(str_section)
53     sections_titles.append(section)
54
55     # Sort sections by _INDEX file.
56     if INDEX_FILE in sections:
57         child_path = path + "/" + INDEX_FILE
58         titles = read(child_path).split("\n")
59         orders = {}
60         level = 1
61         for title in titles:
62             orders[title] = level
63             level += 1
64         if "*" not in list(orders.keys()):
65             orders["*"] = level
66         section_levels = [
67             orders[title] if title in list(orders.keys()) else orders["*"]
68             for title in sections_titles
69         ]
70         str_sections = [x for _,x in sorted(zip(section_levels,
str_sections))]
71     else:
72         str_sections = sorted(str_sections, key=len, reverse=True)
73         content += NEW_LINE(str_sections)
74     return content
75     return travel(args.path)
76
77
78
79 def create_by_file(args) -> str:
80     content = md2tex(read(args.path), level=0)
81     return content
82
83
84 def get_args() -> argparse.Namespace:

```

```

85     logger.info("Parse arguments.")
86     parser = argparse.ArgumentParser()
87     parser.add_argument(
88         "--path", help="The folder/file path for the book", required=True
89     )
90     parser.add_argument(
91         "--bg", help="Background image for the title page", required=False
92     )
93     parser.add_argument(
94         "--simple", help="Output a simple standalone pdf", required=False,
95         action="store_true"
96     )
97     parser.add_argument(
98         "--name", help="The name of the book", required=True
99     )
100    parser.add_argument(
101        "--output", help="The output folder for the pdf and log files, default
102        ='log'", required=False
103    )
104    parser.add_argument(
105        "--author", help="The author of the book", required=False
106    )
107    return parser.parse_args()
108
109 if __name__ == "__main__":
110     args = get_args()
111     logger.info(args)
112     if args.output:
113         output = f"{args.output}/{args.name}.tex"
114     else:
115         output = f"log/{args.name}.tex"
116     output_content = ""
117     if isdir(args.path):
118         output_content = create_by_folder(args)
119     elif isfile(args.path):
120         output_content = create_by_file(args)

```

```

120
121     write(
122         output,
123         DOCUMENT(
124             args.name,
125             output_content,
126             bg_image=args.bg,
127             book_author=args.author,
128             is_simple=args.simple,
129         )
130     )

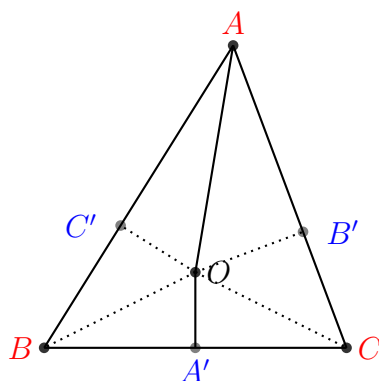
```

## 2.3 导入 tex 格式的图片文件

```

1 {{images/趣题集/三角形悖论/image1.tex}}[image] # 导入tex格式的图片文件

```

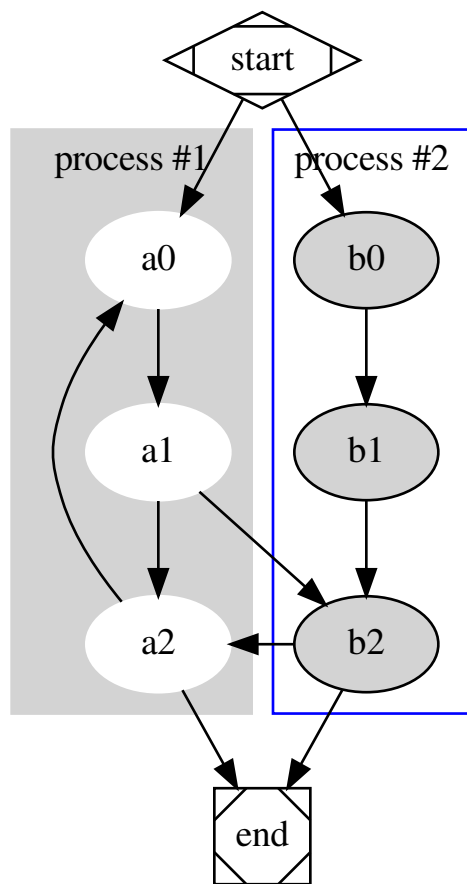


## 2.4 导入 dot 格式的图片文件

注意：需要安装[Graphviz](#)。

导入的 dot 文件会通过 dot 命令转化为 eps 格式，并作为图片导入到 tex 文件中。dot 的简单用法可以参照[GraphViz Documentation](#)。

```
1 {{images/README/image0.dot}}[dot:测试dot] # 导入dot格式/graphviz语法的图片文件，并用“测试dot”作为标题。
```



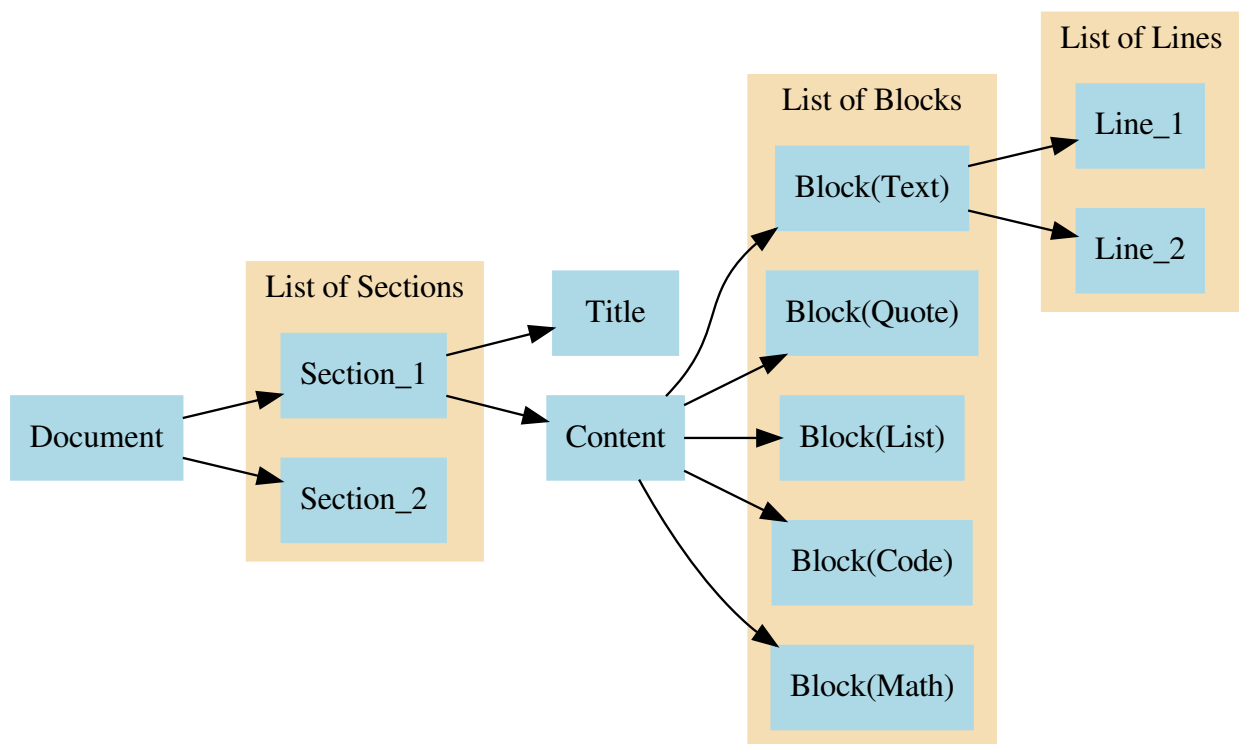
测试 dot

## 2.5 Markdown 的解析和渲染

该系统自带一个 Markdown 解析器，会将 Markdown 文本转化为一棵语法树。使用如下命令可以打印出本 Readme 文档的语法树：

```
1 make test # python3 test.py --path README.md
```

Markdown 语法树的解析遵循如图的树状结构。从最顶端的 Document 到最底端的 Line，而 inline 部分的语法则较为简单，将由渲染器进行解析。



Markdown 语法树

通过渲染器, 可以将 Markdown 渲染成 tex, html, word 等格式。只需要继承 MDRender, 实现对应的抽象方法即可添加新的渲染器, 比如 src/render/md\_tex\_render.py。

```

1 #!/usr/bin/python
2 # -*- coding: UTF-8 -*-
3 from typing import List
4 from .constants import ListType, InLineType, IncludeType
5 import abc
6 import re
7

```

```

8
9 # use (.*?) to extract the pattern
10 # use (?:) to make the extraction optional
11 # use (.*?) to make a non-greedy match (shortest match)
12 # use (.*?)? to make the match optional
13 class MDRender():
14     def __init__(self, level: int = 0):
15         self.level = level
16         self.render_line_configs = {
17             InLineType.Image: {
18                 'pattern_str': r"!\\[(.*?)\\]\\((.*?)\\)",
19                 'group_num': 2,
20             },
21             InLineType.Link: {
22                 'pattern_str': r"\\[(.*?)\\]\\((.*?)\\)",
23                 'group_num': 2,
24             },
25             InLineType.Bold: {
26                 'pattern_str': r"*\\*(.*?)\\*",
27                 'group_num': 1,
28             },
29             InLineType.Include: {
30                 'pattern_str': r"\\{\\[(.*?)\\]\\}\\[(.*?)\\]",
31                 'group_num': 2,
32             },
33         }
34
35     @abc.abstractmethod
36     def render_title(self, title, level) -> str:
37         pass
38
39     @abc.abstractmethod
40     def render_list(self, items: List[str], type_: ListType=ListType.Normal)
41     -> str:
42         pass
43
44     @abc.abstractmethod

```

```

44     def build_inline_link(self, title: str, link: str) -> str:
45         pass
46
47     @abc.abstractmethod
48     def build_inline_image(self, title: str, link: str) -> str:
49         pass
50
51     @abc.abstractmethod
52     def build_inline_bold(self, content: str) -> str:
53         pass
54
55     @abc.abstractmethod
56     def build_inline_include(self, path: str, type_: IncludeType) -> str:
57         pass
58
59     def build_inline_new_pattern(self, match, inline_type: InLineType) -> str:
60         # Image must be before Link, otherwise Image will be treat as Link
because the format
61         # Image: ![]()
62         # Link: []()
63         if inline_type == InLineType.Image:
64             title = match.group(1)
65             link = match.group(2)
66             return self.build_inline_image(title, link)
67         elif inline_type == InLineType.Link:
68             title = match.group(1)
69             link = match.group(2)
70             return self.build_inline_link(title, link)
71         elif inline_type == InLineType.Bold:
72             content = match.group(1)
73             return self.build_inline_bold(content)
74         elif inline_type == InLineType.Include:
75             path = match.group(1)
76             setting = match.group(2)
77             if len(setting.split(":", 1)) == 2:
78                 type_ = setting.split(":", 1)[0]
79                 config = setting.split(":", 1)[1]

```



```

80         else:
81             type_ = setting
82             config = ""
83             return self.build_inline_include(path, type_, config)
84         return ""
85
86     def render_line_with(self, line, inline_type, config) -> str:
87         changed = True
88         while changed:
89             pattern = re.compile(config['pattern_str'], re.IGNORECASE)
90             match = pattern.search(line)
91             if match and len(match.groups()) == config['group_num']:
92                 old_pattern = match.group(0)
93                 new_pattern = self.build_inline_new_pattern(match, inline_type
94 )
95                 line = line.replace(old_pattern, new_pattern)
96             else:
97                 changed = False
98         return line
99
100     @abc.abstractmethod
101     def render_line(self, line: str) -> str:
102         for inline_type, config in self.render_line_configs.items():
103             line = self.render_line_with(line, inline_type, config)
104         return line
105
106     @abc.abstractmethod
107     def render_blockquote(self, content) -> str:
108         pass
109
110     @abc.abstractmethod
111     def render_math(self, content) -> str:
112         pass
113
114     @abc.abstractmethod
115     def render_code(self, content, language) -> str:
116         pass

```

## 第三章 TODO

- 目前很多 latex 设置都是 Hard Code 的，应当将这些设置变成可配置的：
  - 目前目录最多显示两层标题结构
  - 目前最高层目录默认为 Chapter，设置 level 参数使其可配置化
  - 图片默认居中，可以考虑配置化。
- 目前 Markdown 解析器仍然有许多 Markdown 基本语法并不支持，需要：
  - 支持行内斜体，下划线等。
  - 支持行内代码。
  - [Done] 支持一般图片格式的导入。
  - [Done] 支持嵌套 List。
    - \* 测试 3 层嵌套。
  - 支持表格
- 目前 Latex 渲染器仍然有一些 Latex 用法支持不够，需要：
  - 支持更多类似  $\$$ ， $<$ ， $>$  等 latex 特殊字符。
- 支持更多的 Features：
  - 使用 folder 模式的时候，兼容.md 等后缀。
  - 添加页面背景水印，这样可以防止发布的材料被盗版使用。
  - 添加程序辅助画图（特别是制作复杂的各类数学图片）。
  - 添加作者介绍。
  - 添加 html 和 word 渲染器。
  - 支持 latex 画思维导图。
- BUG:
  - tex 图片嵌入应该像 dot 图片嵌入一样支持设置标题，如 [image:title]。
- Better Engineering:

- md\_tree 的 Parser 主体可以用自动机的方式实现，从而减少 if/else 的判断和重复代码片。



扫码支持作者