

Python与Word

create by Dcount

- Python与Word
 - create by Dcount
- 一、简介
 - 1. 使用模块Python-docx
 - 2. word文档结构
- 二、提取文字
 - 1. 读取文档
- 三、写入内容
 - 1. 添加标题
 - 2. 添加段落
 - 3. 添加分页
 - 4. 添加图片
 - 5. 添加表格
 - 6. 保存
- 四、调整样式
 - 1. 文字样式
 - 2. 段落样式
 - 2.1 对齐方式
 - 2.2 行间距
 - 2.3 段间距

一、简介

1. 使用模块Python-docx

导入时使用`import docx`

2. word文档结构

1. Document:文档
2. Paragraph:段落
3. Run:文字块

二、提取文字

1. 读取文档

```
from docx import Document
doc = Document('filename.docx')
```

- `.text` 读取文字内容
- `doc.paragraphs`: 读取文段
- `doc.paragraphs[0].text`: 读取段落文字
- `run = doc.paragraphs[1].runs.text` 提取文字块RUN

三、写入内容

1. 添加标题

`doc.add_heading('name', level)`, 参数为标题名字和标题等级

2. 添加段落

`para1=doc.add_paragraph('内容')` `para.add_run('文字块').bold = True`: 添加字后加粗
`para.add_run('文字块').italic = True`: 斜体

3. 添加分页

`doc.add_page_break()`

4. 添加图片

`doc.add_picture(图片地址, width=doc.shared.Cm(5), height = doc.shared.Cm(5))` 只有给出一个尺寸, 其他的可以自动调整

5. 添加表格

`doc.add_table(rows, cols)` 几行几列的表格 代码示例, 假设有数据records

```
table = doc.add_table(rows=4, cols=3)
for row in range(4):
    cells = table.row[row].cells
    for col in range(3):
        cell[col].text = str(record[row][col])
```

6. 保存

`doc.save(文件路径)`

生成一个请假条

四、调整样式

1. 文字样式

`run.font` 里面调整

```
from docx import Document
from docx.shared import Pt, RGBColor
from docx.oxml.ns import qn
doc = Document( '这是一个文档. docx')
for paragraph in doc.paragraphs :
    for run in paragraph.runs:
        run.font.bold = True#加粗
        run.font.italic = True#倾斜
        run.font.underline= True#下划线
        run.font.strike = True#删除线
        run.font.shadow = True
        run.font.size = Pt(20)#字体大小
        run.font.color.rgb = RGBColor(255,255, 0)#颜色
        run.font.name= '楷体'
        r = run._element.rpr.rFonts
        r.set(qn('w:eastAsia'), '楷体')
doc.save(filename)
```

2. 段落样式

2.1 对齐方式

`paragraph.alignment=对齐方式`，其中对齐方式是在`docx.enum.text.WD_ALIGN_PARAGRAPH`中选取，是枚举类型 有以下几个：

- LEFT: 左
- CENTER: 居中
- RIGHT: 靠右
- JUSTIFY:
- DISTRIBUTE:
- JUSTIFY_MED:
- JUSTIFY_HI:
- JUSTIFY_LOW
- THAI_JUSTIFY

2.2 行间距

`paragraph.paragraph_format.line_spacing = 2.0`:注意用浮点数就表示两倍行间距

2.3 段间距

- `paragraph.paragraph_format.space_before=Pt(12)`,表示段前12磅
- `paragraph.paragraph_format.space_after=Pt(12)`,表示段后12磅