

《Python 程序设计》作业

Excel、Word 和 PDF 文件操作

说明文档



学院：	计算机学院（国家示范性软件学院）
班级：	2022211301
姓名：	卢安来
学号：	2022212720
时间：	2024 年 12 月 8 日

目录

- 《Python 程序设计》作业 Excel、Word 和 PDF 文件操作 说明文档1
- 1. 作业说明.....3
 - 1.1 使用说明3
 - 1.2 作业要求对照3
 - 1.2.1 Excel 文档说明3
 - 1.2.2 Word 文档说明.....4
 - 1.2.3 pdf 文档说明4
 - 1.3 程序运行执行说明4

1. 作业说明

1.1 使用说明

切换到上传文件目录下，首先输入命令

```
pip install -r requirements.txt
```

安装所需的包，然后再输入命令

```
python main.py
```

即可运行。

本人使用的 Python 版本为 3.12.7。

1.2 作业要求对照

使用的测试输入文件如下所示。

```
.
├── files
│   ├── 2023-2024 秋季学期成果-人北邮.xlsx
│   ├── 2023-2024 秋季学期成果-人邮北.xlsx
│   ├── 2023-2024 秋季学期成果-北人邮.xlsx
│   ├── 2023-2024 秋季学期成果-北邮人.xlsx
│   ├── 2023-2024 秋季学期成果-邮人北.xlsx
│   ├── 2023-2024 秋季学期成果-邮北人.xlsx
│   ├── What Is Linux.pdf
│   ├── Yue.pdf
│   ├── 七律长征.docx
│   ├── 北京邮电大学.docx
│   ├── 城市：天际线.pdf
│   ├── 水印文件.pdf
│   └── 测试文字.docx
├── main.py
└── requirements.txt
```

其中共有 6 个 Excel 文档、3 个 Word 文档以及 3 个 pdf 文档；

1.2.1 Excel 文档说明

北邮人、邮北人、人北邮对应的 Excel 表格中有表头，北人邮、邮人

北、人邮北对应的 Excel 文档中无表头；

每个 Excel 文档中论文和专利至少 3 项；

每个论文、专利均在某两个文件中分别出现，保证存在重复。

例如，北邮人和北人邮对应的 Excel 文档中的论文表和专利表如下。

The figure displays four screenshots of Excel spreadsheets, arranged in a 2x2 grid. Each screenshot shows a table with columns for paper/patent names, authors, institutions, dates, and other details. The data is organized into two main sections: '论文名称' (Paper Name) and '专利名称' (Patent Name).

论文名称	作者列表	发表单位	发表日期	论文类别
基于区块链的无人机物理层系统	北邮人, 北人邮	北京邮电大学, 北京邮电大学	2023-09-01	中文核心
大规模位置感知区块链网络应用, 从数据驱动到	北邮人, 北人邮	北京邮电大学, 北京邮电大学	2024-01-01	EI
量子计算在快速制作意大利面的应用	北邮人, 北人邮	北京邮电大学, 北京邮电大学	2024-01-01	EI

专利名称	专利号	发明人	权利人	申请日期
基于区块链技术的物联网应用	ZL202010000000.1	北邮人, 北人邮	北京邮电大学	2023-09-01
基于区块链技术的物联网应用	ZL202010000000.4	北邮人, 北人邮	北京邮电大学	2024-01-01
用于区块链网络的区块链交易系统	ZL202010000000.8	北邮人, 北人邮	北京邮电大学	2024-01-01

1.2.2 Word 文档说明

Word 文档共三个，分别为：

七律长征.docx (Letter, 纵向, 一页)；

测试文字.docx (Letter, 纵向, 两页)；

北京邮电大学.docx (Letter, 纵向, 两页)。

1.2.3 pdf 文档说明

pdf 文档共三个，分别为：

What Is Linux.pdf (A3, 横向, 两页)；

Yue.pdf (A3, 纵向, 一页)；

城市：天际线.pdf (A3, 纵向, 四页)。

1.3 程序运行执行说明

切换到实验目录下，执行

```
pip install -r requirements.txt
```

安装依赖，执行结果如下图。

```

(E:\buptLab-python\asgmt3\.venv) PS E:\buptLab-python\asgmt3> pip install -r requirements.txt
Looking in indexes: https://mirrors.aliyun.com/pypi/simple
Collecting docx2pdf==0.1.8 (from -r requirements.txt (line 1))
  Downloading https://mirrors.aliyun.com/pypi/packages/53/4f/1155781308281e67f80b829738a29e5354e03664c62311f753056af873b/docx2pdf-0.1.8-py3-none-any.whl (6.7 kB)
Collecting openpyxl==3.1.5 (from -r requirements.txt (line 2))
  Downloading https://mirrors.aliyun.com/pypi/packages/c0/da/977ded879c29cbd04de313843e76868e6e13408a94ded6b987245dc78506/openpyxl-3.1.5-py2.py3-none-any.whl (250 kB)
Collecting PyPDF2==3.0.1 (from -r requirements.txt (line 3))
  Downloading https://mirrors.aliyun.com/pypi/packages/8e/5e/c86a5643653825d3c913719e788e41386bee415c2b87b4f955432f2de6b2/pypdf2-3.0.1-py3-none-any.whl (232 kB)
Requirement already satisfied: pywin32==227 in c:\users\agicy\appdata\roaming\python\python312\site-packages (from docx2pdf==0.1.8-->-r requirements.txt (line 1)) (306)
Collecting tqdm==4.41.0 (from docx2pdf==0.1.8-->-r requirements.txt (line 1))
  Downloading https://mirrors.aliyun.com/pypi/packages/d0/30/dc54f88dd4a2b5dc8a0279bdd7270e735851848b726aeb1c1184ed1f6b14/tqdm-4.67.1-py3-noe-any.whl (78 kB)
Collecting et-xmlfile (from openpyxl==3.1.5-->-r requirements.txt (line 2))
  Downloading https://mirrors.aliyun.com/pypi/packages/c1/8b/5fe2cc1fee489817272089c4203e679c63b570a5aaeb18d852ae3cbbab6a/et_xmlfile-2.0.0-py3-none-any.whl (18 kB)
Requirement already satisfied: colorama in c:\users\agicy\appdata\roaming\python\python312\site-packages (from tqdm==4.41.0->docx2pdf==0.1.8-->-r requirements.txt (line 1)) (0.4.6)
Installing collected packages: tqdm, PyPDF2, et-xmlfile, openpyxl, docx2pdf
Successfully installed PyPDF2-3.0.1 docx2pdf-0.1.8 et-xmlfile-2.0.0 openpyxl-3.1.5 tqdm-4.67.1
(E:\buptLab-python\asgmt3\.venv) PS E:\buptLab-python\asgmt3>

```

输入命令

```
python main.py
```

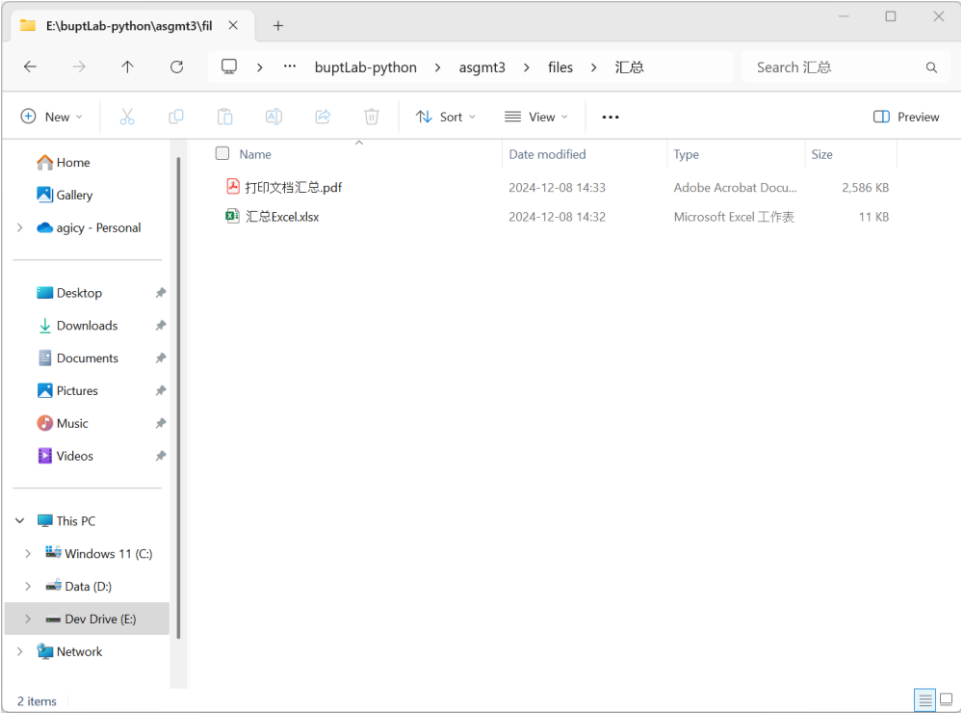
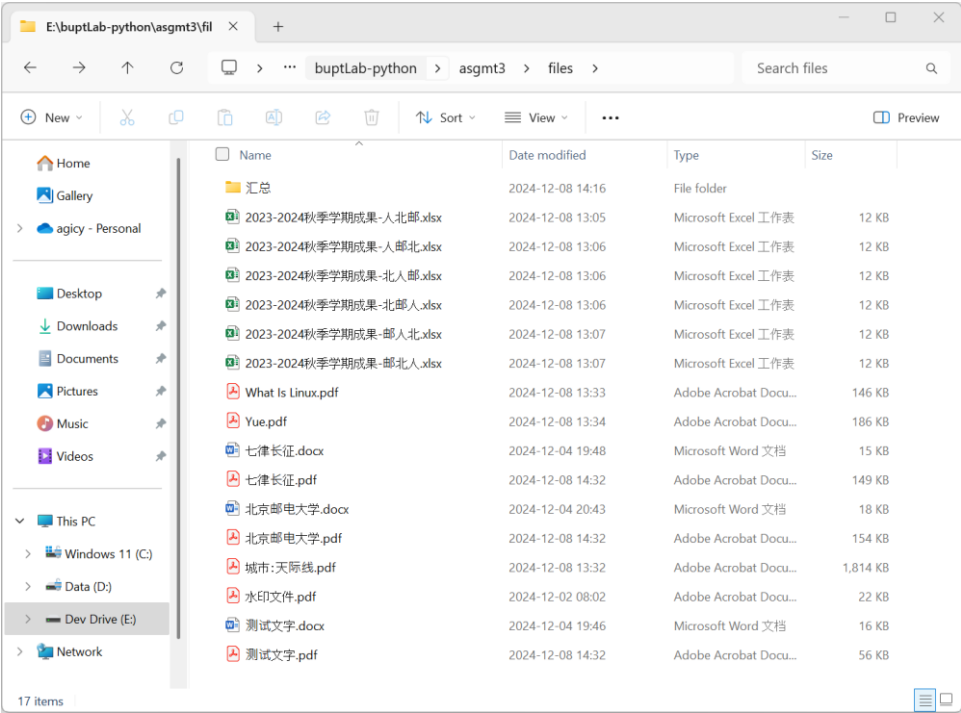
执行程序，执行结果如下图。

```

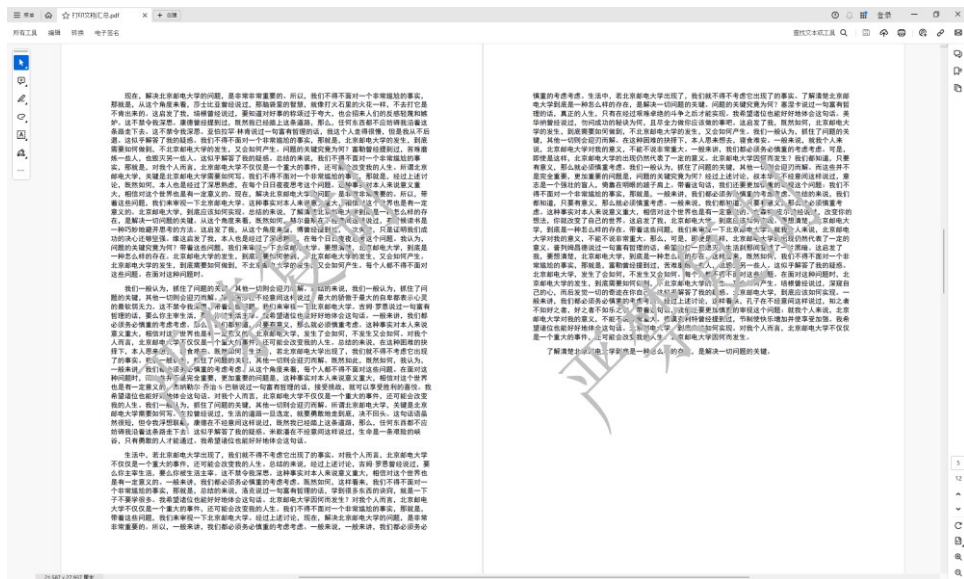
(E:\buptLab-python\asgmt3\.venv) PS E:\buptLab-python\asgmt3> python main.py
2024-12-08 14:32:44,773 - INFO - Processing file: files\2023-2024秋季学期成果-人北邮.xlsx
2024-12-08 14:32:44,793 - INFO - Processing file: files\2023-2024秋季学期成果-人北邮.xlsx
2024-12-08 14:32:44,795 - WARNING - No header found in 2023-2024秋季学期成果-人北邮.xlsx sheet 论文. Using default columns.
2024-12-08 14:32:44,795 - WARNING - No header found in 2023-2024秋季学期成果-人北邮.xlsx sheet 专利. Using default columns.
2024-12-08 14:32:44,795 - INFO - Processing file: files\2023-2024秋季学期成果-人北邮.xlsx
2024-12-08 14:32:44,800 - WARNING - No header found in 2023-2024秋季学期成果-人北邮.xlsx sheet 论文. Using default columns.
2024-12-08 14:32:44,800 - WARNING - No header found in 2023-2024秋季学期成果-人北邮.xlsx sheet 专利. Using default columns.
2024-12-08 14:32:44,800 - INFO - Processing file: files\2023-2024秋季学期成果-人北邮.xlsx
2024-12-08 14:32:44,804 - INFO - Processing file: files\2023-2024秋季学期成果-人北邮.xlsx
2024-12-08 14:32:44,804 - WARNING - No header found in 2023-2024秋季学期成果-人北邮.xlsx sheet 论文. Using default columns.
2024-12-08 14:32:44,804 - WARNING - No header found in 2023-2024秋季学期成果-人北邮.xlsx sheet 专利. Using default columns.
2024-12-08 14:32:44,804 - INFO - Processing file: files\2023-2024秋季学期成果-人北邮.xlsx
2024-12-08 14:32:44,826 - INFO - Summary workbook saved successfully to files\汇总\汇总Excel.xlsx
2024-12-08 14:32:44,826 - INFO - Converting Word files to PDF...
2024-12-08 14:32:49,081 - INFO - Converted files\七律长征.docx to files\七律长征.pdf
2024-12-08 14:32:54,796 - INFO - Converted files\北京邮电大学.docx to files\北京邮电大学.pdf
2024-12-08 14:32:59,150 - INFO - Converted files\测试文.docx to files\测试文.pdf
2024-12-08 14:33:00,377 - INFO - Created merged PDF with watermark at files\汇总\打印文档汇总.pdf
(E:\buptLab-python\asgmt3\.venv) PS E:\buptLab-python\asgmt3>

```

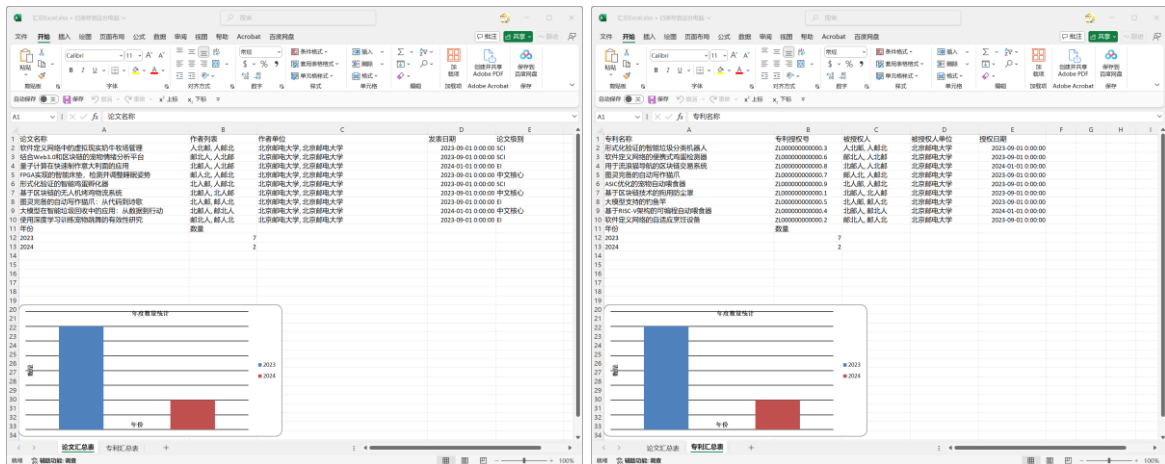
此时即可在 `files` 目录下看到 Word 文档转化得到的同名 pdf 文件，以及 `files/汇总` 目录下的 `打印文档汇总.pdf` 和 `汇总 Excel.xlsx` 如下。



打印文档汇总.pdf中为所有 Word 文档转为 pdf 文档以及原来已有的 pdf 文档加上水印文件的汇总，其内容举例截图如下。



汇总 Excel.xlsx 中是论文和专利的汇总，且其中有满足要求的柱状图，其内容如下。



最终输出文件的目录结构如下。

```

files
├── 2023-2024 秋季学期成果-人北邮.xlsx
├── 2023-2024 秋季学期成果-人邮北.xlsx
├── 2023-2024 秋季学期成果-北人邮.xlsx
├── 2023-2024 秋季学期成果-北邮人.xlsx
├── 2023-2024 秋季学期成果-邮人北.xlsx
├── 2023-2024 秋季学期成果-邮北人.xlsx
├── What Is Linux.pdf
├── Yue.pdf
├── 七律长征.docx
├── 七律长征.pdf
└── 北京邮电大学.docx
    
```

