2019-2020学年 第 一 学期

**实验报告**

课程名称 Python程序设计实验

系（部） 信息工程系

专业班级 电子信息科学与技术18-2

学生姓名 段明宇

学生学号 201823010205

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 实验名称 | | 正则表达式与文件内容操作练习 | | | | |
| 实验目的 | | 熟练掌握正则表达式的使用和对文件操作的内容 | | | | |
| 实验内容 | | 第八章、第九章的练习题，对9-17反向应用。 | | | | |
| 实验仪器设备 | | Jupyer Notebook、AI Studio、电脑 | | | | |
| 姓 名 | | 段明宇 | | 组 别 |  |
| 同组实验者 | |  | | 实验日期 | 2019.10.09 |
| 指导教师 | | 魏光村 | |  |  |
| 批阅意见  及 成 绩 | | |  | | | |
| **实**  **验**  **步**  **骤** | 1 实验原理  2 实验过程记录  3 实验结果  4实验过程中存在的问题及解决方案  5 实验总结 | | | | | |

1 实验原理

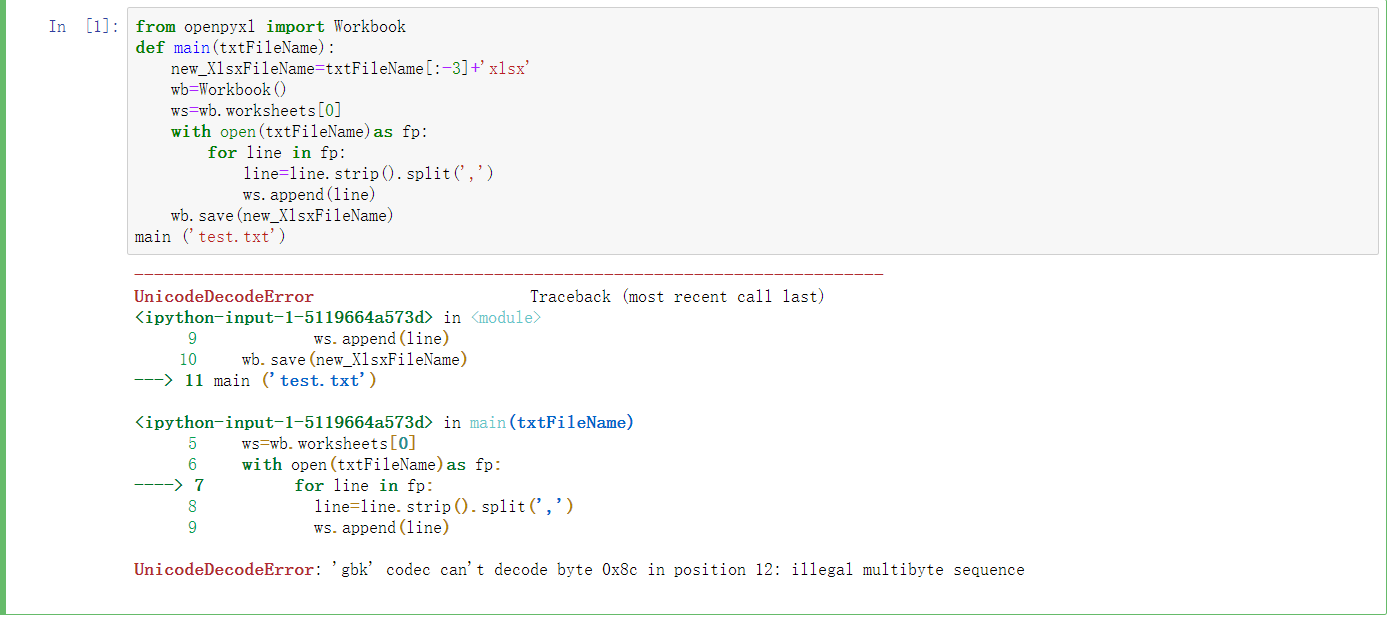
文件是长久保存信息并允许重复使用和反复修改的重要方式，同时也是信息交换的重要途径。记事本文件、日志文件、各种配置文件、数据库文件、图像文件、音频和视频文件、可执行文件、Office文件、动态链接文件等，都以不同的文件形式储存在各种存储设备中。按数据的组织形式，可以把文本分为文本文件和二进制文件两大类。

2 实验过程记录

先对9-17进行情景再现，以更熟悉操作机制。

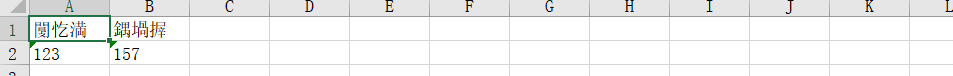


调节后的问题入下：



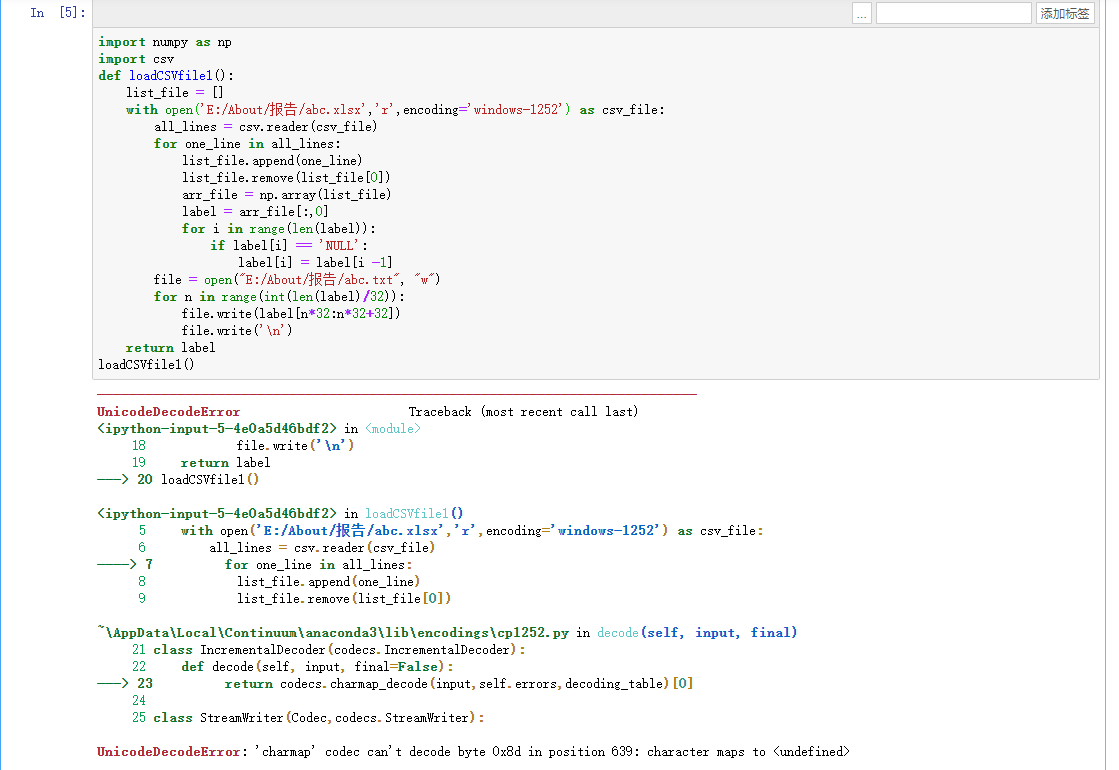
对test.txt进行转换，转换前后如下图所示：





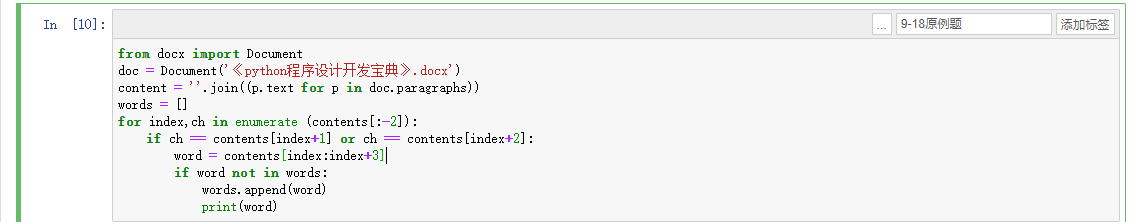
之后进行Excel 2007+文件转换成记事本文件test.txt。在8-17程序中将txt与xlsx交换位置，进行试验。开始思路为将程序中的txt全部与xlsx调换，但转换后出来的文件是xtxt类型，且内容为乱码。因此，改变思路。

在处理数据时需要读取TXT文件类型的数据转换为可执行的list，而在反向转换时则需要import必要的两个包，然后读取Excel文件，然后进行一次额外的筛选，筛选的目的是因为我们所采集到的数据中一般都有一些数据缺失的情况，也就是显示为“NULL”的情况，或者其他的形式，可以根据自己数据的缺失情况进行灵活更改（如果没有数据缺失这部分代码就不需要）之后便是储存到指定的TXT文件中。实验程序过程如下：



经过不断查询与努力思考，最后还有的问题就是对二进制的转化问题。程序的改善有待交流……

第二个问题则是检查Word文档的连续重复字，如“用户的的资料”或“需要需要用户输入”之类的情况以及各种word中的各种错误的查找与改善。本例使用扩展库python-docx读写Word 2007+文档的内容。



而后，其他的问题如多余空格、多余标点等错误，均可用此代码进行转用。

3 实验结果

9-17例题能准确运行，并且能够解决实际问题。而在反向应用时则遇到各种问题，直到最后还存在的问题，如字节的转换，之后继续思考此问题。另一个问题则较为简单与方便。直接能够运行并利用。

4实验过程中存在的问题及解决方案

4.1搜索不到文件；

4.2不会调用包；

4.3不知道如何读取过大文件；

解决方案：换地址创建新的文件，准确的途径打开文件。文件过大转用其他方式。

5 实验总结

文件是长久保存信息并允许重复使用和反复修改，同时也是信息交换的重要途径。

二进制文件无法使用记事本直接打开并正常阅读和修改，必须使用相应的软件进行有关操作。

文本文件和二进制为念的操作流程是一样的。

进行文件内容的读写操作时推荐使用上下文管理语句with。