

## 20181024 作业

- 1、作业压缩包中有一个名为“My\_Fifteen\_Lost\_Years.txt”的英文小说文件。完成如下要求：
- 2、原始数据处理。将文件中所有非英文字母(指 A-Z、a-z 之间的英文字母)包括换行在内一律替换为空格。后续所有题目都在此条件下完成；
- 3、请将所有的大写字母转换为小写字母；
- 4、统计每个字符出现的次数从小到大排列，并利用该数据形成 Excel 折线图，将该折线图存为图形文件为学号\_cf.png。字母频率保存到一个文件(学号\_cf.txt)，第一列为字母，第二列字符出现次数，两列之间用英文逗号间隔；
- 5、统计每个单词(转换为小写之后，连续字母且长度大于等于 2)中字母的间距，并计算出每个字母之间的平均间距，比如：ab 间距、ac 间距等等，统计相同字母之间的间距，比如：bee 中 e 与 e 之间的间距。如在 word 之中，wo 之间的间距为 1，wr 的间距 2，w 和 d 的间距为 3；o 与 r 的间距为 1，o 与 d 的间距 2，r 与 d 的间距为 1，以此类推。字间距按从高到低的顺序保存到文件中，名为学号\_cd.txt。第一列是类似 ab、ac 的两个字符组合，第二列为字间距，两列之间用英文逗号间隔；
- 6、（思考题）根据本文中的英文小说的字母频率以及字间距，评估当前键盘的键位分布是否合理。思考键盘布局应考虑哪些要素等等。此小题不提交作业文件，但要求确实进行思考；
- 7、作业提交内容：源代码(命名为学号\_cp.py)、学号\_cf.txt、学号\_cd.txt、学号\_cf.png；上述两个 txt 文件不允许手动先创建一个空文件，而是用 Python 打开一个不存在的文件然后写入数据。Python 在打开一个文件用于写时，会自动创建一个文件以接受输入；
- 8、作业提交截止日期为：2018 年 10 月 30 日 24:00 前，本次作业不能延期提交。