

20181031 作业

- 1、 压缩文件中的 `isChinese.py` 程序中有一个 `isChinese()` 函数，用于判断一个字符是否是汉字。利用该函数，将“红楼梦”和“水浒传”中的非汉字清除；
- 2、 分别统计红楼梦和水浒传中每个汉字出现的频率（出现次数除以总字数，以清除非汉字字符后计算），分别按序输出频率最高的 500 个汉字及其频率，**频率保留 8 位小数**。汉字一列，频率一列，中间用英文逗号隔开，将结果输出到 `2.txt` 中。
- 3、 按顺序输出红楼梦中汉字个数，水浒传中汉字个数，只存在于红楼梦中的汉字个数，只存在于水浒传中的汉字个数。一行输出，中间用英文逗号隔开，将结果输出到 `3.txt` 中。
- 4、 第一行输出只在红楼梦中出现的汉字；第二行输出只在水浒传中出现的汉字，汉字之间用英文逗号隔开。列出时，按从高到低顺序排列，保存到 `4.txt` 中；
- 5、 以红楼梦字的频率减去水浒传对应字的频率的绝对值从大到小排序，输出排序最前的 500 个字（此题仅计算红楼梦和水浒传共有字）；输出一行，汉字之间用英文逗号隔开，保存到 `5.txt` 中。
- 6、 用第 5 问的数据在 Excel 中做柱状统计图，分别生成前 25 字、前 50 字、前 100 字的柱状统计图。将统计图存为 PNG 格式图片，保存为文件，分别命名为 `top25.png`、`top50.png`、`top100.png`；
- 7、 作业提交内容：源代码和三张统计图；不需要提交生成的 `txt` 文件，同意压缩成 `zip` 压缩包提交，`zip` 命名为：学号.`zip`；要求 1 到 5 题的源代码必须在同一 `python` 文件中，并且生成的 `txt` 文件均以相对路径形式保存于源代码同一路径下。
- 8、 此次助教会同时给出 `windows` 和 `mac` 下的判断程序，要求大家的代码严格通过判断程序，判断程序之后会另行

给出。

9、 作业提交截止时间：2018-11-13 23:59 前。