

## 20181017 作业

- 1、 作业文件中有一个名为“50W.txt”的文件，编码格式为 utf-8，内容为近 50 万人名，每个名字一行。编程时可能用到字符串的 split()函数，请参看讲义“快速入门”字符串 str 部分的讲解；
- 2、 利用上述数据，统计每个姓出现的次数，不考虑复姓问题(复姓视为单姓)，将排名靠前的 100 个姓以降序输出到文件，格式形如：王,1000;李,900;张,880。以此类推；
- 3、 利用上述数据，统计最后一个字出现的次数，将排名靠前的 100 个字以降序输出到文件，格式形如：丹,1000;兵,900;国,880。以此类推；
- 4、 利用上述数据，统计最后两个字中每个字出现的次数，将排名靠前的 100 个字以降序输出到文件，格式形如：国,1000;蓉,900;建,880。以此类推；
- 5、 利用上述数据，统计 3 个及 3 个以上名称前两个字(整体)出现的次数，将排名靠前的 20 个以降序输出到文件，格式形如：欧阳,100;南宫,80;王国,60。以此类推；
- 6、 利用上述数据，统计两个字(整体)连续相同出现的次数(间隔相同不算，比如：王大王，虽有连个王，但不算，目标是统计：李蓉蓉、王丹丹中的蓉蓉、丹丹)，将排名靠前的 20 个以降序输出到文件，格式形如：丽丽,100;莉莉,80;丹丹,60。以此类推；
- 7、 以上第 2-6 小题，每个小题输出为一行，全部输出到文件名为“学号\_20181017.txt”之中，其中学号替换为自己的学号；
- 8、 程序文件命名为“学号\_20181017.py”；
- 9、 利用“学号\_20181017.txt”数据在 Excel 形成数据列表和统计图(图形类型为折线图)。建议用 Word 完成查找替换等，然后将数据拷贝到 Excel 并成图。Excel 文件命名为“学号\_20181017.xlsx”；
- 10、 提交内容：数据文件学号\_20181017.txt、程序文件学号\_20181017.py、Excel 文件学号\_20181017.xlsx；
- 11、 作业提交时间：2017-10-23 24:00 前。