

20200226 作业

- 1、 作业文件中有一个名为“50 万人名.txt”的文件，编码格式为 utf-8。
- 2、 按姓名字数排序，即输出 2 字姓名、3 字姓名、4 字姓名的数量；格式为：第一列分别为 2 字、3 字、4 字；第二列为对应的数量；列与列之间用\t 分隔（制表符，阅读讲义）；
- 3、 统计每个姓出现的次数，不考虑复姓问题(复姓视为单姓，第一个字为姓)，按降序排列。第一列为姓的汉字，第二列为对应数量，列与列之间用\t 分隔；
- 4、 统计名(姓之外的字)中每个字出现次数，相当于统计名称常用字，不考虑复姓问题(视为单姓)，按降序排列；第一列为汉字，第二列为对应数量，列与列之间用\t 分隔，仅输出前 200 个汉字；
- 5、 统计名(第一个字为姓)中有两字重复出现的次数，相当于“丽丽”出现的次数，“萌萌”出现的次数。如果是四个字，如：欧阳阳春，阳阳算重复，欧阳丽丽，丽丽算，花花美，花花不算。第一列为重复的两个字，第二列为对应的次数，列与列之间用\t 间隔；
- 6、 提交内容包括：程序源代码(第 2-5 问，每个 1 个程序，程序命名为:学号_编号_Work01.py)；产生的数据(命名规则如程序，扩展名为 txt)，程序源码注释丰富，变量命名规范。将上述内容压缩到学号_Work01.zip 中；
- 7、 作业完成完成 2020-03-04 18:00 前。