

## 20190419 作业

- 1、 下载链接：“<https://pan.baidu.com/s/1Fun5YhreAdSnUGbTUCiqVw> 提取码：ryev”处的数据(数据较大，3.6GB，下载时要有耐心)，实现如下功能；
- 2、 编写一个名为“hzFollow.py”的 Python 程序，计算出每个汉字后跟随的不超过 20 个高频汉字，计算结果保存到文件 hzFollow.txt 之中，每一行的首字为被计算汉字，从第二个开始为跟随汉字及其频次，如：高□兴,100;文,80;大,60;寒,30;【注：□是半角空格】
- 3、 数据规模较大，注意策略。比如：拆分数据，用一部分数据调试程序等；
- 4、 【附加，不计入评分标准，完成且比较科学可增分，不完成不扣分】20 个高频汉字属于生硬规定，感兴趣的同学，可根据字频分布规律制定策略限制跟随汉字的数量，即有些汉字跟随多，有些跟随少，如：“中”的跟随就会多，“彬”跟随比较少；
- 5、 编写一个名为“hzInput.py”的 Python Flask 程序，利用 hzFollow.txt 实现联想式汉字输入法，即在网页上每输入一个汉字，在网页按字频显示出跟随的汉字，且能实现按编号(0-9)选择跟随汉字；
- 6、 网页端显示效果自定，美观大方有助于得分；
- 7、 作业完成完成 2019-05-15 18:00 前。