6.6 文本词频统计

问题分析：

- 需求：一篇文章，出现了哪些词语？哪些词语出现得最多？

英文还是中文？

-英文文本：Hamet 分词词频

<https://python123.io/resources/pye/hamlet.txt>

-中文文本:《三国演义》 分词人物

<https://python123.io/resources/pye/threekingdoms.txt>

针对于：文本噪音处理，归一化，提取每个单词为第一个步骤，统计每个单词出现得词频数量

《三国演义》人物出场统计

将词频与人物相关联，面向问题

从词频统计到人物统计

# 举一反三

应用问题的扩展

* 其他名著
* 政府工作报告等，词频分析
* 进一步，未来还有词云