

## 14. 排序

在文件 `a-tale-of-two-cities.txt` <sup>1</sup>中，保存着著名小说家 Charles Dickens 的小说《A Tale of Two Cities》(公元 1859 年)。假定单词的定义是“仅由字母组成的最长的连续的字符串”。以下问题忽略大小写。

1. 统计文档中有多少个单词。
2. 统计文档中有多少个不同的单词。
3. 打印出现次数最多的前 20 个单词以及它们的词频。

---

<sup>1</sup> 见古登堡计划的网站. <http://www.gutenberg.org/ebooks/98>