

计算概论

作业题

2021.11.23

作业题

- 场景

- 统计英文文学作品中的词频

- 具体要求

- 选择一篇自己熟悉的英文文学作品
 - 读入文本，进行预处理
 - 用nltk库分词和去除停用词
 - 增设自定义的停用词列表，并去除自定义停用词
 - 将词频最高的10个词输出在控制台窗口（print输出）
 - 将出现超过3次的词和词频写入CSV文件
 - 用Excel打开CSV文件，选择合适的图表进行展现

- 提交要求

- 12月5日提交，包括（1）源代码文件（包含详细的注释），（2）输入文件（txt格式的英文文学作品），（3）CSV结果文件（包含数据和图表）