

《Python 程序设计高阶》

(2023-2024 学年第 1 学期)

作业 2

学号:2021329600006 姓名: 陈昊天 班级: 计算机科学与技术 21(4) 班
学号:2021329621213 姓名: 陈佳伟 班级: 计算机科学与技术 21(3) 班
学号:2021329621257 姓名: 冯佳钧 班级: 计算机科学与技术 21(4) 班

§1 题目 1

作者鉴定

现在我们的目光转向古代，一直有人认为《水浒传》的作者不是施耐庵，而是罗贯中。
现在就请你对《水浒传》来做一个判定，看看和《三国演义》到底风格是否相近。
鉴定方法由你自己来设计，当然，必须用 Python 来实现哦。
看看最后你的结论是什么？

§1.1 解答

通过统计两书虚词 [1] 所占比例比较其相似度。

1.1.1 数据来源

水浒传: <https://shuihu.5000yan.com/xiazai/>

三国演义: <https://sanguo.5000yan.com/xiazai/>

```
1 def count_function_words(file_path):
2     with open(file_path, "r", encoding="utf-8") as file:
3         text = file.read()
4
5     # 虚词列表
6     function_words = [
7         "之", "其", "或", "亦", "方", "于", "即", "皆", "因", "仍", "故", "尚",
8         "呢", "了", "不", "乃", "呀", "吗", "啊", "把", "让", "向", "往",
9         "在", "比", "好", "可", "便", "就", "还", "依", "只", "犹", "才",
10        "早", "没", "吧", "过", "得", "地", "个", "著", "上", "着", "起",
11        "来", "下", "去", "进", "回", "北", "南", "东", "西", "头", "里",
12        "外", "前", "后", "内", "中", "间"
13    ]
14
15    total_words = len(text)
16    function_words_count = 0
17
18    for word in function_words:
```

```

14         function_words_count += text.count(word)
15
16     function_words_ratio = function_words_count / total_words
17
18     return function_words_ratio
19
20
21 # 分别统计《水浒传》和《三国演义》的虚词比例
22 shuihu_ratio = count_function_words("shuihu.txt")
23 sanguo_ratio = count_function_words("sanguo.txt")
24
25 # 输出结果
26 print("《水浒传》虚词比例:", shuihu_ratio)
27 print("《三国演义》虚词比例:", sanguo_ratio)
28
29 # 比较虚词比例，判断风格是否相近
30 if abs(shuihu_ratio - sanguo_ratio) < 0.1:
31     print("《水浒传》和《三国演义》风格相近")
32 else:
33     print("《水浒传》和《三国演义》风格不相近")

```

```

● (base) nanmener@Haotians-MacBook-Pro python % /Applications/Ana
conda3/anaconda3/bin/python "/Users/nanmener/Github/zstu-study/
Python程序设计-高阶/Python programming@ZSTU by Laitude/4. Lists
, Tuples, Sets & Dictionaries/style.py"
《水浒传》虚词比例：0.13426335870847347
《三国演义》虚词比例：0.10753506775748008
《水浒传》和《三国演义》风格相近

```

§2 总结

§2.1 分工情况

以下说明小组的分工情况，每个人完成了什么功能，完成的情况如何

陈昊天

负责报告的整合撰写；

编写文件读取的代码；

陈佳伟

搜集、完善虚词列表；

冯佳钧

编写统计、比较虚词比例的代码；

§2.2 解决方案评价

以下说明小组对解决方案的评价，是否完成了题目的要求，怎么证明完成了题目的要求，好的地方在哪里，哪里有不足。

基本完成了题目的要求；通过统计虚词比例比较两文文风相似度；

优点：代码结构清晰明确；时间复杂度为 $O(\text{len}(\text{虚词列表}) * \text{len}(\text{书籍长度}))$ ；

缺点：仅统计了虚词比例，没有使用其他的文言词汇，结果可能不准确；

参考文献

- [1] Wikipedia contributors. Function word — Wikipedia, the free encyclopedia, 2023. [Online; accessed 26-October-2023].