

基于python 的豆瓣网站书籍数据爬取分析与可视化 系统的设计与实现

目 录

第 1 章 绪论.....	1
1.1 研究背景及意义	1
1.2 国内外研究现状	1
1.3 论文研究及设计内容	2
第 2 章 设计所用相关技术.....	3
2.1 爬虫部分.....	3
2.1.1 Requests.....	3
2.1.2 Beautiful Soup.....	3
2.1.3 Pandas.....	3
2.1.4 Time.....	4
2.2 数据分析与可视化部分.....	4
2.2.1 matplotlib.....	4

2.2.2 Seaborn.....	4
2.2.3 Pandas.....	5
2.3 前端展示部分.....	5
2.3.1 pyecharts.....	5
第3章 网络爬虫.....	6
3.1 数据来源.....	6
3.2 过程.....	6
3.2.1 网页板块分析.....	6
3.2.2 网页元素审查.....	7
3.2.3 爬虫代码编写.....	8
3.3 获取数据结果.....	10
第4章 数据分析与数据可视化.....	11
4.1 数据分析.....	11
4.2 数据可视化.....	14
4.3 结果分析.....	18
第5章 前端可视化展示.....	19
5.1 数据分析并制作可视化图像.....	19
5.2 可视化大屏生成.....	27
第6章 结语.....	29
参考文献.....	30
致谢.....	31

摘要:

在数据爆炸式增长的现代化社会，各类信息十分繁杂，一个统计分析并清晰展现数据的网页将对读者筛选图书具有较强的参考价值。本研究将利用 Python 编程语言对豆瓣网站书籍 Top 250 部分进行网络爬虫，获取相关书籍信息。随后采用 pandas 库进行数据处理，确保数据的准确性和完整性。基于处理后的数据，运用 matplotlib 库进行数据可视化，以直观方式展示书籍的关键指标如排名、评分等，从而进行进一步的数据分析。最后，利用 pyecharts 库制作六类不同的交互式可视化图像，并将其转化为 HTML 前端网页。本研究为用户提供了获取书籍相关信息的平台，便于用户浏览与交互，进而深化对豆瓣书籍 Top 250 排行榜的理解，在数据的基础上筛选目标书籍。

关键词: Python 爬虫；数据分析；数据可视化

第 1 章 绪论

1.1 研究背景及意义

在科技高度发展的现代社会，大数据无处不在，各类信息呈爆炸式增长，人们在快节奏的生活中，时间与神经经常处于紧绷状态，在这样的状况中，许多人会选择通过阅读书籍来进行调节，在书的世界中寻求心灵的放松。但挑选一部自己心仪的书籍并不是十分容易的事，往往需要查询作者、读者评价等多种信息，互联网中的信息虽已十分完备，但各类排行榜中的书籍排名并不适用于所有读者的喜好，查询的过程使读者还没有开始阅读已经耗费大量时间及精力，这个过程极易降低读者的阅读意愿。因此，若能够将书籍的数据整合并清晰地展现出来，帮助读者直观迅速地观察数据、理解数据之间的联系，将大大提升人们的阅读意愿，更加轻松地选择适合自己喜好的书籍，享受阅读带来的乐趣。

1.2 国内外研究现状

互联网已成为现代人获取信息的首选渠道，使用便捷成为了人们使用互联网的最大理由。而爬虫技术为互联网更添一份便利，通过爬虫，用户可以迅速、准确地获取到所需资源，爬虫的应用领域十分广泛，其中最流行的便是爬取网页^[1]。Python 作为近年来十分常用且强大的编程语言，为爬虫技术提供了许多

方法和工具，不仅能够高效地抓取信息，也能在避免反爬虫方面发挥重要作用，确保用户能够安全、稳定地获取所需数据。结合 Python 语言的独特优势与互联网未来的蓬勃发展趋势，可以得知 Python 爬虫技术是一项具备广阔拓展空间的前沿技术，掌握爬虫技术知识不仅对个人成长大有裨益，更是能契合社会和时代对科技进步与创新的不懈追求^[2]。

数据可视化技术的发展最早可以追溯到 18 世纪前后应用的手工测量绘图及统计学图表^[3]，如今，随着计算机技术的发展，数据可视化技术的形式与范围也愈发完善，逐渐从平面二维向三维立体、动态交互发展。经过查阅发现，数据可视化的研究在国内外已十分普遍，在金融、医疗^[4]、教育及城市发展等多个领域均扮演着举足轻重的角色，使用数据可视化是对数据进行分析展示的十分适配的方法。

1.3 论文研究及设计内容

因此，本研究将利用 Python 爬虫技术，对市场主流网站豆瓣网中的书籍 Top250 排行榜进行精确的数据提取。获取到数据后，将借助数据分析及可视化工具如 pandas 和 seaborn 等，对这些数据进行深入剖析。通过科学的分析，可以揭示出书籍排行榜背后的规律与趋势，为读者提供更为精准科学的筛选方法。基于这些分析结论，利用数据可视化绘图工具 pyecharts，绘制出书籍信息相关的交互式图像，直观地展示书籍的各项数据，并令读者通过交互操作，深入了解书籍的数据情况。最后将把这些图像组合制作为清晰美观的可视化大屏网页，将书籍数据一目了然地呈现在读者面前，帮助他们更加快捷便利地挑选到符合预期的心仪书籍。

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：

<https://d.book118.com/335032301212012024>