

大学生论文检测系统

文本复制检测报告单 (全文标明引文)

№: ADBD2023R_2023041311242620230415120334470711805905

检测时间: 2023-04-15 12:03:34

篇名: 计算机科学与技术本科1909班_张永远_180508010832_基于Python的知乎话题数据可视化系统设计与实现

作者: 张永远

指导教师:

检测机构: 郑州工商学院

文件名: 计算机科学与技术本科1909班_张永远_180508010832_基于Python的知乎话题数据可视化系统设计与实现.docx

检测系统: 大学生论文检测系统

检测类型: 大学生论文

检测范围: 中国学术期刊网络出版总库

中国博士学位论文全文数据库/中国优秀硕士学位论文全文数据库

中国重要会议论文全文数据库

中国重要报纸全文数据库

中国专利全文数据库

图书资源

优先出版文献库

大学生论文联合比对库

互联网资源(包含贴吧等论坛资源)

英文数据库(涵盖期刊、博硕、会议的英文数据以及德国Springer、英国Taylor&Francis 期刊数据库等)

港澳台学术文献库

互联网文档资源

源代码库

CNKI大成编客-原创作品库

机构自建比对库

时间范围: 1900-01-01至2023-04-15

检测结果

去除本人文献复制比: 5.8%

跨语言检测结果: 0%

去除引用文献复制比: 2.2%

总文字复制比: 5.8%

单篇最大文字复制比: 1.7% (医疗贴吧中广告的提取系统)

重复字数: [942]

总段落数: [2]

总字数: [16338]

疑似段落数: [2]

单篇最大重复字数: [273]

前部重合字数: [88]

疑似段落最大重合字数: [756]

后部重合字数: [854]

疑似段落最小重合字数: [186]



文字复制部分 2.2%

引用部分 3.6%

无问题部分 94.2%

指标: ☐ 疑似剽窃观点 ☒ 疑似剽窃文字表述 ☐ 疑似整体剽窃 ☐ 过度引用

相似表格: 0

相似公式: 没有公式

疑似文字的图片: 0

7.5% (756)

7.5% (756)

计算机科学与技术本科1909班_张永远_180508010832_基于Python的知乎话题数据可视化系统设计与实现_第1部分 (总10083字)

3% (186)

3% (186)

计算机科学与技术本科1909班_张永远_180508010832_基于Python的知乎话题数据可视化系统设计与实现_第2部分 (总6255字)

指导教师审查结果

指导教师:

审阅结果:

审阅意见: 指导老师未填写审阅意见

1. 计算机科学与技术本科1909班_张永远_180508010832_基于Python的知乎话题数据可视化系统设计与实现_第1部分 总字数: 10083

相似文献列表

去除本人文献复制比: 7.5%(756) 文字复制比: 7.5%(756) 疑似剽窃观点: (0)

1	<u>医疗贴吧中广告的提取系统</u> 张园园(导师: 冯洪海) - 《河南大学硕士论文》 - 2016-06-01	2.7% (273) 是否引证: 是
2	<u>基于大数据技术的电信客户维系挽的分析与研究</u> 潘娜(导师: 徐江峰) - 《郑州大学硕士论文》 - 2017-05-01	1.5% (149) 是否引证: 否
3	<u>基于大数据的图书阅读分析平台</u> 王利祥 - 《大学生论文联合比对库》 - 2020-06-01	1.3% (127) 是否引证: 否
4	<u>基于大数据的图书阅读分析平台</u> 王利祥 - 《大学生论文联合比对库》 - 2020-06-09	1.3% (127) 是否引证: 否
5	<u>20167711238-王利祥-基于大数据的图书阅读分析平台的设计与实现</u> 王利祥 - 《大学生论文联合比对库》 - 2020-06-13	1.3% (127) 是否引证: 否
6	<u>基于Python的文本情感分析模型设计</u> 曾聪 - 《大学生论文联合比对库》 - 2019-05-12	0.8% (78) 是否引证: 否
7	<u>基于自然语言的电影评论分析</u> 林媛媛 - 《大学生论文联合比对库》 - 2020-03-30	0.8% (78) 是否引证: 否
8	<u>3150931022-郑天方-基于python的网络爬虫设计及课程评论分析系统开发-张亚玲</u> 郑天方 - 《大学生论文联合比对库》 - 2019-06-03	0.7% (72) 是否引证: 否
9	<u>202004100250585254_刘燕_红色电影评论自动采集与分析方法研究</u> 刘燕 - 《大学生论文联合比对库》 - 2020-04-10	0.7% (72) 是否引证: 否
10	<u>202004100250585254_刘燕_红色电影评论自动采集与分析方法研究</u> 刘燕 - 《大学生论文联合比对库》 - 2020-04-10	0.7% (72) 是否引证: 否
11	<u>基于Python的三种网络爬虫技术研究</u> 杨健;陈伟; - 《软件工程》 - 2023-02-05	0.7% (68) 是否引证: 是
12	<u>豆瓣电影评论文本的情感分析及主题提取研究</u> 余洋(导师: 石磊) - 《云南财经大学硕士论文》 - 2018-09-10	0.6% (59) 是否引证: 是
13	<u>项目管理平台的设计与实现</u> 张成强;卢凤娟; - 《电脑知识与技术》 - 2021-05-25	0.5% (46) 是否引证: 是
14	<u>基于APICloud的齿轮箱监控平台移动端研究</u> 吴鲁纪;耿福震;杨林杰;李俞峰; - 《电脑知识与技术》 - 2019-11-25	0.5% (46) 是否引证: 否
15	<u>基于多源数据的铁路出行乘客用户画像及方案推荐研究</u> 吴运腾(导师: 杨敏;汪林) - 《东南大学硕士论文》 - 2020-05-01	0.5% (46) 是否引证: 否
16	<u>现制茶饮大数据报表-凌霞-专接本毕业设计</u> 凌霞 - 《大学生论文联合比对库》 - 2021-04-23	0.4% (42) 是否引证: 否
17	<u>基于知识图谱的JAVA程序课程答疑设计与应用研究</u> 伍思洁(导师: 顾倩颐) - 《四川师范大学硕士论文》 - 2019-03-20	0.4% (40) 是否引证: 否

原文内容

本科毕业设计
基于Python的知乎话题
数据可视化系统设计与实现
院部名称信息工程学院
姓名张永远
学号 180508010832
专业计算机科学与技术
届别 2023届
指导教师李敏
2023年4月21日
基于Python的知乎话题数据可视化系统设计与实现
摘要：随着现在物

院部名称	信息工程学院
姓名	张永远
学号	180508010832
专业	计算机科学与技术
届别	2023届
指导教师	李敏

质生活的逐步提高，人们对生活、知识和思想的探究深度也在提高。知乎是目前国内非常受欢迎的问答社区，用户通过回答别人的问题或者回复回答来分享知识资源或者通过浏览和讨论来汲取所需的资源。本文通过分析从知乎爬虫得到的问题和对应浏览量、话题等数据进行热点分析，分析热点可以分析社会发展、了解时事、分析舆论走向。数据分析通过可视化可以直观且清晰的发现规律，发现热搜、热榜会带来巨大流量，可以用来获得更多的信息。本文主要工作有后端通过Python和爬虫来进行数据收集，所需技术和库有bs4、re正则表达、urllib、SQLite存储数据、time。可视化阶段使用Flask搭建网站，使用Echarts制作图表、WorldCloud制作词云来实现可视化效果，用到的其他包有jieba、matplotlib、numpy。通过本系统完成了对各指定知乎话题精华问题的爬取，使用此数据做出了观察浏览量与回答数量的散点图，根据数据中的问题或话题做出了词云图。散点图可以看出对于大多数话题中回答量和浏览量的正相关性是很明显的。词云图在看出学科话题下可推论知乎用户的学历和年龄结构是大学及以上有很高比例。通过分析数据提升了数据敏感度，提高了多维度看待问题的能力。

关键字：python；网络爬虫；数据分析；数据可视化

Design and Implementation of Zhihu Topic Data Visualization System Based on Python Crawler

Abstract: With the gradual improvement of material life, the depth of people's exploration of life, knowledge and thought is also increasing. Zhihu is currently a very popular question-and-answer community in China. Users share knowledge resources by answering other people's questions or replying to answers, or obtain the required resources by browsing and discussing. This article conducts hotspot analysis by analyzing the questions obtained from the Zhihu crawler and the corresponding pageviews, topics and other data. Analyzing hotspots can analyze social development, understand current events, and analyze public opinion trends. Data analysis can intuitively and clearly discover the rules through visualization, and discover that hot searches and hot lists will bring huge traffic, which can be used to obtain more information. The main work of this article is to collect data through python and crawlers at the back end. The required technologies and libraries include bs4, re regular expression, urllib, SQLite to store data, and time. In the visualization stage, flask was used to build the website, ECharts was used to make charts, and WorldCloud was used to make word clouds to achieve visualization effects. Other packages used were jieba, matplotlib, and numpy. Through this system, the crawling of the essence questions of each designated Zhihu topic is completed. Using this data, a scatter diagram of the observed number of views and the number of answers is made, and a word cloud diagram is made based on the questions or topics in the data. From the scatter plot, it can be seen that the positive correlation between the number of answers and the number of views in most topics is obvious. The word cloud map can be inferred from the subject topics that Zhihu users have a high proportion of college and above in their education and age structure. By analyzing the data, the data sensitivity is improved, and the ability to look at problems from multiple dimensions is improved.

Keyword: Python; Web Crawler; Data Analysis; Data Visualization

目录

1 绪论	6
1.1 研究目的和意义	6
1.2 国内外文献综述	7
1.3 研究的主要内容和方法	8
1.3.1 研究内容	8
1.3.2 研究方法	8
1.4 系统技术介绍	9
1.4.1 Python爬虫介绍	9
1.4.2 Flask框架介绍	9
1.4.3 Echarts图表介绍	10

1.4.4 WordCloud词云介绍.....	10
2系统分析	11
2.1 可行性分析.....	11
2.1.1 经济可行性分析.....	11
2.1.2 技术可行性分析.....	11
2.1.3 操作可行性分析.....	12
2.2 需求分析.....	12
2.2.1 功能需求分析.....	12
2.2.2 性能需求分析.....	13
3 系统总体设计	14
3.1 爬虫功能设计.....	14
3.2 数据库设计.....	14
3.3 数据可视化设计.....	15
3.3.1 Flask框架设计.....	15
3.3.2 Echarts图表设计.....	16
3.3.3 WordCloud词云设计.....	16
4 系统实现	17
4.1 系统开发的平台和运行环境.....	17
4.1.1 系统开发环境平台.....	17
4.1.2 运行所需包.....	17
4.2 后台功能的实现.....	17
4.2.1 爬虫功能的实现.....	17
4.2.2 Flask功能实现.....	20
4.2.3 Echarts图表功能实现.....	20
4.2.4 WordCloud词云功能实现.....	21
4.3 系统前台界面.....	21
4.3.1 主页功能界面.....	22
4.3.2 数据展示界面.....	22
4.3.3 Echarts图表界面.....	23
4.3.4 WordCloud词云界面.....	25
4.4 数据库的连接.....	27
4.4.1 爬虫部分数据库连接.....	27
4.4.2 可视化部分与数据库连接.....	27
5 系统测试	28
5.1 测试系统的目的.....	28
5.2 测试方法.....	28
5.3 测试结果.....	30
结论	31
参考文献	32
致谢	33

1 绪论

1.1 研究目的和意义

我国一直注重经济发展，但对文化产业的支持也在增加。知乎是我国最受欢迎的知识问答社区。用户通过共享信息资源，了解自己需要的资源。在付费内容领域，知乎月活跃付费用户数已超过250万，总内容数超过300万，年访问人次超过30亿[1]。知乎的核心作用是了解用户提出的问题匹配其他用户推荐其回答问题。在知乎上，用户对提出的问题进行标记其所属话题，系统也会通过标记找到相关问题推荐给用户。网络爬虫的出现对于提升搜索的覆盖率和精准率有着很重要的意义[2]。

如今，无论是哪个行业，良好的数据分析都非常重要。信息提取是指从特定的信息流中将人们感兴趣的信息过滤出来，在本文中的信息提取可以转化为文本的分类问题[3]。数据分析只是对特定数据的准确分析。数据分析的目的是集中、提炼和改进隐藏在大量无组织数据中的信息，以发现研究成分的内在规律。

事实上，数据分析可以帮助人们做出能做什么的决定。数据分析也是规划和收集数据、分析目的地并将其转化为信息的过程。数据分析具体可以：分类、预测分析、关联规则和推荐系统、数据缩减和降维、数据探索和可视化。这些都是后台统计分析数据的结果，商人赚的就是信息差，信息差的来源之一就是数据，分析数据可以获得更多的信息。

总的来说数据分析的意义就是告诉过去发生了什么，这些现状为什么会发生，以及未来会发生什么。

数据可视化是将数据转换为图形或图表的过程，以便更轻松地理解和分析数据。数据可视化的意义在于：更好地理解数据，通过图形或图表的形式展示数据，更容易发现数据中的规律和趋势。发现数据中的问题和机会，通过图表和图形的分析，可以发现某些数据点的异常情况，从而发现需要解决的问题或利用的机会。提高决策效率，通过图表和图形的展示，可以迅速地了解数据，以便更快地做出决策。提高交流效率，数据可视化可以更好地与他人交流。通过图表和图形的展示，可以更清晰地表达数据和信息，从而更好地与他人沟通和合作。

各种个性化推荐让人们束缚在信息茧房中，本文可以提一种以宏观角度分析信息热点的功能，让人们不只是被迫接受算法推荐的信息，更全面以多角度看待问题。大数据时代人们需要更强的整理信息和多维度分析信息的能力，人没有交流讨论就没有思想碰撞就不能有进步，知乎可以说部分代替了旺盛时期的贴吧、论坛。有一个自动收集整理信息并将信息进行可视化供分

析的工具具有非常重要的意义。

1.2 国内外文献综述

潘娜在《基于大数据技术的电信客户维系挽留的分析与研究》使用Python语言进行网络爬虫的研究实现，爬取新浪微博的大量数据。搭建了分布式系统。在数据可视化方面，使用Python的Matplotlib绘图库进行可视化图表分析。依据关键词搜索进行了二次筛选，提取出用户发布的与电信企业相关的数据内容，方便下一步对筛选后的数据进行多维度的可视化分析。作者将处理和分析后的数据存储在MongoDB中，然后使用Matplotlib绘图库进行可视化分析[2]。

张圆圆在《医疗贴吧中广告的提取系统》中介绍了Python和爬虫，编写了爬虫程序，实现了贴吧中文本数据的抓取。使用有 Python 的 Numpy、Matplotlib 绘图工具。对训练样本进行分词和去停用词：分词采用jieba分词工具实现，停用词表采用网上开源词表，针对本课题特点，对停用词表进行修正[3]。

孙立伟等在《网络爬虫技术的研究》使用了Python中的多个库来实现数据爬取和多线程管理，具体包括但不限于：requests、BeautifulSoup等。与传统手工数据采集相比，该系统有以下优势：自动化程度高，可以对数据进行自动处理和分析，可以实现定时更新数据，可以避免人为因素对数据采集过程的影响[4]。

余洋在《豆瓣电影评论文本的情感分析及主题提取研究》使用Python网络爬虫技术抓取了豆瓣电影Top250的评论文本数据作为研究对象。在此基础上，对文本数据进行了清洗和预处理。然后对评论信息进行初步处理。然后对处理过的信息进行情感上的判断和分析，主要是在评论中提取出带有情感特征的名词，匹配情感词典，设计出一种打分机制来计算情感综合值[5]。

梁融凌等在《基于Python爬虫的胡润百富榜数据可视化分析》使用了 Python 网络爬虫和数据分析技术并介绍了详细步骤，使用了 Python 的可视化库 Matplotlib 和 Seaborn 来对数据进行可视化展示。这些图表可以直观地展示不同富豪的财富分布、行业分布、地域分布等信息，并且可以通过交互式操作进行筛选和排序[6]。

Yu L等在《Summary of web crawler technology research》文章中阐述了网络爬虫在搜索引擎上扮演着重要的角色，尤其是在提取网页时。Web爬虫最重要的作用是在Internet的大数据中爬行，查找有效的信息，并将所需的信息数据存储在本地数据库中，是穿越超链接和索引的计算机程序[7]。

1.3 研究的主要内容和方法

1.3.1 研究内容

熟悉网页页面结构，正则表达式、bs4网页解析提取页面元素，运用爬虫库requests框架、数据库增删改查，能根据需求，处理常见的反爬，抓取数据。

用Flask进行网页搭建、Echarts对数据进行可视化处理，WordCloud进行绘制词语图，了解聚类分析等分析方法。

熟悉使用html web开发。能对常见数据载体格式进行数据的解析。

数据分析是一种常用的统计方法，其主要功能是多维和描述性的。一些几何技术显示了不同数据之间的关系，并提取统计信息以更简洁地解释此数据中的重要信息。其他用于收集数据以找出谁是同质的，以便更好地理解数据。

数据分析可以处理大量数据并识别该数据中最有用的部分。近年来该领域的成功可能主要归功于制图技术的改进。这些图可以通过直接分析数据来揭示难以察觉的联系。更重要的是，它与现象分布无关，与经典的统计方法相反。数据分析是数学和计算机科学相结合的结果。数据分析是为了提取有用信息和形成结论而对数据加以详细研究和概括总结的过程[8]。

通过分词技术，可以粗略地看出用户普遍看重的方面。在词云图中，词汇越大，说明该词汇在文本中出现次数越多，越能代表更多的用户[9]。

1.3.2 研究方法

本文涉及的研究方法主要有观察法和数量研究法。

观察法：通过和已学知识进行匹配观察确定研究对象。

数量研究法：通过爬取到的大量数据进行分析揭示事物间的关系、规律和趋势，达到解释和预测。

1.4 系统技术介绍

1.4.1 Python爬虫介绍

本文选择 Python 作为此系统的开发语言。Python开发环境定义了列表、字典、元组等许多高级数据类型。这些是本文中提取数据所需的。与其他编程语言相比，可以简化很多代码并使其更具可读性，包括作为 Python 语言特色之一的正则表达式。使编程更容易的强大功能。Python 语言是可以面向过程或面向对象的编程语言，易于学习。Python 编程环境还提供了一种交互式编程模式，方便用户在开发应用程序时监控和查看应用程序的内容[10][11]。

爬虫是获取、检索数据一种方式，能按照一定规则自动抓取某个站或者万维网信息的程序；现实环境中大部分网络访问都是由爬虫造成的。本文从爬虫和数据处理分析两部分来展开。

网络爬虫又被称为网络蜘蛛，是一段可以自动抓取 Web 上信息的程序或脚本[10]。网络爬虫按照系统结构和实现技术实际应用中通常是将系统几种爬虫技术相互结合[4]。网络爬虫，它的本质是一组计算机程序[12]。要完成对网页的下载、搜索要按指定要求获取相关数据。并且这个过程无需用户干预而循环执行。通过请求页面上的 HTML 文档来识别对特定页面的访问。检查现有站点，不断在站点之间移动，自动创建文件并在内部存储它们网络数据库。当网络爬虫进入超链接时，它会搜索信息并接收其他超链接。该 URL 主要基于HTML 结构，不依赖于用户干预[3]。首先向待爬取的网站发起请求，如果目标网站的服务器响应正常，会得到响应，然后再通过合理技术手段解析目标网页内容，最后选择爬取、保存所需要的文本数据[5]。

爬虫就是把网页中需要的数据通过指定标签和属性的方式提取出来，然后通过正则表达更准确的提取出标签内的所需数据。

1.4.2 Flask框架介绍

Flask是一个用Python编写的轻量级Web应用程序框架，相对于Django来说更加简洁、灵活和可扩展。通过定制第三方扩展，能够快速实现特定的功能，提高了可定制性和开发效率。同时，Flask也具备处理用户请求和数据传递的能力，在Web开发中广泛应用。

Flask框架包括WSGI工具库Werkzeug和模板引擎Jinja2两个主要的库，具有简单、灵活、可扩展等优点，适用于各种规模的Web应用程序。Flask已经成为越来越多开发人员构建Web应用程序的首选框架，因为它易于使用，同时提供强大的功能。

1.4.3 Echarts图表介绍

Enterprise Charts 简称为 ECharts, 这是一个使用JavaScript 实现的开源可视化库, 涵盖各行业图表[13]。

ECharts缩写来自Enterprise Charts, 商业级数据图表, 可以流畅的运行在PC和移动设备上, 兼容当前绝大部分浏览器[14]。具有各种各样的图表, 丰富的动态效果。在官网有详细的教程和非常多的示例, 能在官网实时渲染修改代码。

Echarts主要用于用于构建交互式的数据可视化。它提供了多种图表类型, 如线图、柱状图、散点图、饼图、地图等, 并支持灵活的配置选项, 使得开发人员可以轻松地创建漂亮、交互式的图表和数据可视化应用程序。Echarts还支持响应式布局, 可以自适应不同的设备屏幕大小。

Echarts还支持多种数据格式, 包括常见的JSON、XML、CSV等, 同时也支持异步加载数据。它提供了强大的数据处理能力, 如数据过滤、排序、聚合等, 使得开发人员可以轻松地处理和展示大量的数据。

除此之外, Echarts还支持多种图表类型和数据格式, 提供强大的数据处理能力和灵活的配置选项。它还支持响应式布局、数据交互和动态更新等特性, 使得开发人员可以快速构建漂亮、交互式的可视化应用程序。

1.4.4 WordCloud词云介绍

WordCloud是一种基于Python的文本可视化工具, 将文本数据转换为词云图形式的图表。它通过对文本数据中每个词出现的频率进行统计和分析, 生成一个可视化的词云图, 每个单词的大小表示该单词在文本中出现的频率, 单词的颜色则可以根据不同的主题或情感进行着色表达。WordCloud支持自定义字体、颜色、形状等参数, 适用于文本数据分析、情感分析、主题分析等领域, 也可以用于制作海报、PPT等设计。

2 系统分析

2.1 可行性分析

2.1.1 经济可行性分析

爬虫数据可视化市场需求大, 获取数据成本低。数据分析和可视化过程投入的成本也较低, 包括算法开发、工具选型和可视化设计等方面。商业模式和盈利渠道包括出售可视化报告、提供数据分析咨询服务和提供数据可视化工具。

在生产和生活中获取更多的信息有非常大的帮助, 而把信息以一种简单且直观的方式展现更是如虎添翼。人工时间成本升高, 为减轻重复性劳动负担, 爬虫数据量之大不能使用人工方式代替。开发、维护此系统的成本较少, 因此在经济上是可行的。

2.1.2 技术可行性分析

在爬虫数据可视化项目中, 技术可行性至关重要。需要使用多种技术完成数据采集、清洗、存储、处理和可视化等多个环节。具体来说:

(1) 数据采集: Python的网络爬虫库可以用来爬取知乎网页上的数据。确定数据来源并选择适合的爬虫框架进行采集。同时需要考虑网站反爬机制对采集的影响及应对方法。

(2) 数据清洗: 清洗和去重数据需要使用相关的工具和算法, 例如正则表达式、BeautifulSoup、Pandas等。

(3) 数据存储: 通常使用数据库进行存储, 需要选择适合的数据库类型和对应库, 例如MySQL、MongoDB等。

(4) 数据处理: Python拥有丰富的数据处理和分析库, 可以对获取到的知乎数据进行处理和清洗, 这些库提供了各种数据处理和分析的功能。使用相关的数据处理算法和工具对原始数据进行处理和分析, 以得到更有价值的信息。

(5) 数据可视化: Python拥有众多的数据可视化库, 可以用于生成丰富多样的可视化图表可以满足不同需求的可视化需求。使用合适的可视化库和工具(例如Echarts、Matplotlib、WordCloud等)展示处理后的数据。

本系统采用Python语言使用爬虫技术, 可视化分析采用Flask框架、Echarts制图技术和WordCloud词云技术。后台数据库采用轻量化关系数据库管理系统SQLite, 本系统在技术方面是可行的。

2.1.3 操作可行性分析

(1) 爬虫数据获取: 爬虫需要访问网络并获取数据, 需要应对反爬机制并选择合适的数据存储方式和格式。

(2) 数据清洗和处理: 爬虫获取的数据需要进行清洗和处理, 包括去重、筛选有效数据、数据类型转换等。在处理时需要考虑算法复杂度、数据量大小对时间和资源的影响。

(3) 可视化技术选择: 选择合适的可视化技术是关键。要根据数据特点和分析目的来选择合适的图表类型和展示方式。需要考虑技术复杂度、学习难度以及技术的稳定性和扩展性。

(4) 系统设计和开发: 系统的设计和开发需要考虑用户体验和可用性。要设计简洁明了、易于操作的界面, 并提供直观、清晰的数据展示和分析功能。

本系统用户的操作只是在网页来观看信息展示, 展示信息获取处理是在后台处理, 用户不能干预。网页操作简单, 风格、效果展示简介易懂, 因此本系统在用户操作上也是可行的。

2.2 需求分析

2.2.1 功能需求分析

本系统主要是实现爬虫及数据可视化系统

图2-1 总体思路概括

图2-2 可视化概述图

2.2.2 性能需求分析

(1) 爬虫对于性能要求不高。

(2) 本文要爬虫的内容不是很巨大, 对于数据库选择SQLite足够满足使用。

(3) Flask是轻量级WEB框架, Echarts是百度开源框架, 能够在绝大多数计算机上流畅运行, WordCloud如果设置生成图片的dpi高些会导致生成速度不高, 不过可以在后端提前生成好, 对用户体验不影响。

3 系统总体设计

使用爬虫系统来进行数据获取, 爬取的数据使用可视化系统展现。包括数据获取、数据处理与分析、数据可视化、界面设计、系统交互等模块, 通过这些模块的协作, 实现对知乎话题数据的可视化展示分析。

3.1 爬虫功能设计

知乎问题爬虫的具体项目有: 问题、话题、浏览量、关注量、回答数量。

图3-1 爬虫的具体项目结构图

3.2 数据库设计

本系统使用了两个数据库，每个数据库有一张表。一个是处理话题精华问题的链接和问题题目，另一个是爬虫的数据。爬虫数据表如表3-1所示。

表3-1 数据储存表

字段名	数据类型	完整性约束	说明
id	int	主键	序号
link	int	非空	链接编码
wenti	varchar	无约束	问题
huati	varchar	无约束	话题
guanzhu	int	无约束	关注量
liulan	int	无约束	浏览量
huida	int	无约束	回答量

字段名数据类型完整性约束说明

id int 主键序号
link int 非空链接编码
wenti varchar 无约束问题
huati varchar 无约束话题
guanzhu int 无约束关注量
liulan int 无约束浏览量
huida int 无约束回答量

下图3-2说明数据的流向，主要是爬虫系统的数据采集过程，可视化系统只涉及到数据的读取。

图3-2 数据流图

3.3 数据可视化设计

3.3.1 Flask框架设计

运行可视化系统即可打开网页查看四个主页面及其包含若干的子页面。在主页下有数据展示、图表、词云和团队链接，在数据展示和图表下有具体话题的链接，词云页面是直接显示所有内容。

图3-3 可视化系统Flask框架

3.3.2 Echarts图表设计

设计关于回答数量和浏览量的散点图，回答量和浏览量大部分都是数量少，因此点局部集中要能够对局部放大缩小。

3.3.3 WordCloud词云设计

由于知乎话题结构就是像一个有根树，所以选择背景图为一颗树的图片。

- (1) 读取数据库内容。
- (2) 借助 jieba 分词库对中文进行分词，去除连接词、停顿词。
- (3) 创建 WordCloud 对象，设置基本属性（背景图、宽、高、字体等）
- (4) 生成词云图，保存或显示图片[6]。

4 系统实现

4.1 系统开发的平台和运行环境

4.1.1 系统开发环境平台

运行环境是Windows10，开发本系统是在PyCharm下进行，PyCharm是主要用于python开发的集成环境。

4.1.2 运行所需包

在爬虫阶段有：bs4用于从 HTML中快速地提取指定的数据。re模块：提供了对正则表达式的支持，进行匹配字符串。urllib操作网页URL，获取网页数据和获取网页异常信息、sqlite3进行SQLite数据库操作、time 在本系统用于设置等待时间。

在数据可视化部分在Flask里导入有flask包、sqlite3，在网页中导入echarts.min.js文件，在WordCloud部分导入有wordcloud，中文分词工具jieba，matplotlib：一个综合库，主要用于绘图，numpy矩阵运算，sqlite3。

4.2 后台功能的实现

4.2.1 爬虫功能的实现

首先分析网页，分析提取网页链接，分析网页具体代码所对应要爬取信息编写对应代码。网站通常由三部分组成：超文本标记语言（HTML）、层叠样式表（CSS）和活动脚本（JScript）。

本系统使用Requests技术，此技术属于爬虫基础性工具包，它模拟人输入网址向服务器递交网络请求，实现自动爬取HTML网页页面信息的功能[15]。

指 标

疑似剽窃文字表述

- 1. 关键词搜索进行了二次筛选，提取出用户发布的与电信企业相关的数据内容，方便下一步对筛选后的数据进行多维度的可视化分析。
- 2. 介绍了详细步骤，使用了 Python 的可视化库 Matplotlib 和 Seaborn 来对数据进行可视化展示。

相似文献列表

去除本人文献复制比：3%(186) 文字复制比：3%(186) 疑似剽窃观点：(0)

1	豆瓣电影评论文本的情感分析及主题提取研究 余洋(导师：石磊) - 《云南财经大学硕士论文》 - 2018-09-10	1.2% (73) 是否引证：是
2	182_肖嘉敏_基于Python和ROSTCM6的在线教育评论数据分析系统 晓红 - 《大学生论文联合比对库》 - 2019-06-09	1.0% (65) 是否引证：否
3	我国对违反“爬虫协议”行为的法律规制研究 曹阳; - 《江苏社会科学》 - 2019-05-10	0.7% (44) 是否引证：是
4	高校办公自动化管理系统设计与实施-论文修改-20181012 - 《大学生论文联合比对库》 - 2018-10-12	0.5% (29) 是否引证：否

原文内容

获取待爬取网页的链接，还需要对每一个url做去重检测。
解析 html 网页源代码获得文本信息，长时间或持续的爬取一个网站可能会触发网站的反爬虫机制，因此需要对爬虫技术做伪装，采用设置（user-agent），可将登录的cookie信息通过user-agent一同存在请求中发给浏览器。

要抓取网页，首先需要分析网页设计。许多网站现在使用一种称为Ajax（异步加载）的技术。当打开一个网站时，会首先看到上面的一些，其余的会慢慢加载。因此，可以查看许多可以滑动的网页，并且某些网页可以在导航时缓慢加载大量信息。这种页面的优点是页面加载速度非常快（因为不必一次加载所有内容）。但是，这项技术不适合爬虫，此时需要花一点力气。知乎不滑动只出现五个回答，由于此次爬虫不爬取评论内容，所以这个问题不需要解决。

第一步：判断爬取的数据是不是静态数据，查看网页的源码。了解基本的HTML格式，使用浏览器的F12功能键打开开发者工具定位到要爬取的内容相关语句，并记录下来在接下来的爬虫过程中要用到。由于大部分网站都有反扒措施，需要一些手段来应对，包括模拟浏览器头部信息，向服务器发送消息，表示告诉服务是什么类型的机器、浏览器（本质上是告诉服务器，可以接受什么类型的文件内容）。具体方法就是在要访问的页面按F12，然后点击Network→刷新页面→马上点击开发者工具的红色点stop，鼠标定位到时间轴最左面→点击name下面一行然后出现Headers就是浏览器发送给服务器的头部内容，在headers最下面“User-Agent”是浏览器标识，要加入urllib里模拟正常浏览器访问，否则是直接发送给浏览器python版本号。“爬虫协议”另一个是Allow或Disallow值，用于设置特定搜索引擎所能访问或禁止访问的具体内容[16]。

图4-1 headers头部信息

还需要导入time模块添加等待时间来防止被反爬。经过尝试设置3秒最合适。

第二步：发送网络请求。导入发送请求的模块requests，打开PyCharm，点击文件菜单→设置→项目→Python解释器，再点击右侧出现的加号，在弹出页面输入要添加的模块，最后点击安装包等待提示安装成功即可。

图4-2 安装第三方包requests

可以按以上步骤安装bs4、re、urllib、SQLite等包。

第三步：创建爬虫首先确定将请求发送到何处以及如何发送。然后使用从requests库请求中获取网页的源代码，用Beautiful Soup分析网站，然后从源代码中查找提取数据。Beautiful Soup已经转移到 bs4 库。这意味着需要在导入BeautifulSoup 之前设置 bs4 库。再使用正则表达抽离出所需信息，最后使用SQLite保存信息。由于时间关系本次只获取根话题即其六个子话题里的问题。

核心代码：

正则表达

rehuati = re.compile('<meta content="(.*?)".*/>') # 题目

rewenti = re.compile('title">(.*?)</h1>') # <h1 class="QuestionHeader-title">马斯克称「人类如果不多生孩子，文明将会崩溃」，如何看待其言论？</h1>

for popover in soup.select("meta[name='keywords']"):

huati = re.findall(rehuati, str(popover))[0] # findall第二个参数需要字符

爬虫结果如下图：

图4-3 根话题数据展示

图4-4 学科话题数据展示

4.2.2 Flask功能实现

使用Flask框架搭建网站，在Flask内编写处理数据传递到网页，编写需要展示的网页，在网页内使用Echarts制作图表Fladk核心代码：

echarts图表

data = cur.execute(sql)

for list in datalist:

lis = []


```
lis.append(int(str(list[6]).replace("None", "0"))) # 部分数据回答量为None 不使用数值转换会报 'int' object has no attribute 'replace'
```

```
return render_template("tubiao.html", data=huidaliulan, name=name)
```

4.2.3 Echarts图表功能实现

在官网有很多案例能清楚的分析代码和图形直接的联系，教程也很详细，能在官网代码编写过程实时查看图形。本章设计使用散点图来观察回答数量与浏览量的关系。

核心代码：

```
series: [  
    {  
        type: 'scatter',  
        data: {{data}},  
        dimensions: ['x', 'y'],  
        symbolSize: 8,  
        itemStyle: {  
            opacity: 0.8  
        },  
    },  
]
```

4.2.4 WordCloud词云功能实现

分词作为一种基本技术，目前由很多较成熟的标记化工具可以使用，例如Stanford Tokenizer、OpenNLPTokenizer、jieba以及哈工大LTP等[17]。本文使用jieba。

停用词来源于网络上整理好的停用词表，删除了停用词的分词词表相对来说更为简洁[18]。

WordCloud核心代码：

绘制图片

```
fig = plt.figure(1)  
plt.imshow(wc)  
plt.axis('off') # 是否显示坐标轴  
plt.savefig(r'./static/assets/img/问题生活、艺术、文化与活动话题.jpg', dpi=800) # 输出词云图片到文件
```

4.3 系统前台界面

系统的界面设计包含了系统的主页、数据展示页面Echarts图表、词云以及团队功能界面的设计与实现。系统界面简洁明了，操作简单，设计首先在所有网页顶部均有通往各个主页面的导航条，通过上方的导航条可随意切换。

4.3.1 主页功能界面

运行Flask后点击下方出现网站即可在浏览器打开主界面，在主页主要是展示其他主页面的简单数据和链接。

图4-8 主页界面

4.3.2 数据展示界面

进入首先是选择具体话题的页面，选择后是以列表的形式显示爬取过的信息。

图4-9 数据展示选择部分

图4-10 数据展示部分子页面

4.3.3 Echarts图表界面

(1) 与数据展示页面类似先进入选择具体问题页面，多了一个全部数据的选项。选择后才可进入查看图表。

图4-11 Echarts图表展示

(2) 图表页面可以用鼠标滚轮和拖动或者下方和右方拖动条对图表进行局部缩放查看。在产业话题中回答量和浏览量的正相关性是整体最明显的，同时也有回答量很多但浏览量不多的问题，可见对于单个问题回答量和浏览量没有绝对相关性。

图4-12 Echarts图表子页面产业话题

图4-13 Echarts图表子页面产业话题放大

(3) 全部数据就是把以上七个话题数据库的叠加。

图4-14 Echarts图表子页面全部数据

图4-15 Echarts图表子页面全部数据放大

在三千浏览量以下虽然回答量和浏览量没有绝对关系，但随着浏览量的增加回答数量比例也在相应增加。

大部分问题都是只有很小的阅读量和浏览量，最大浏览量达七千万，相应的回答数量也有四千七：什么叫降维打击？三体（书籍），「形而上」话题，降维打击。

4.3.4 WordCloud词云界面

由于词云生成需要一定的时间，所以需要在后台提前生成，生成的图片添加到网页上。词云页面采取了直接全部显示后台生成的所有话题的词云。

图4-16 根话题词云

图4-17 学科话题词云

学科话题下问题的关于大学、考研比例很高，并且对比其他话题的问题量此话题的问题量也很高，由此可推断知乎用户比例较大的学历和年龄结构是大学及以上。

4.4 数据库的连接

4.4.1 爬虫部分数据库连接

预处理系统会把知乎精华问题网页的数据处理提取出链接和问题题目，放入预处理数据库，然后爬虫主程序会把预处理数据库中的链接依次读取进行爬虫操作，爬取的数据存入新的爬虫数据库。

4.4.2 可视化部分与数据库连接

可视化系统只涉及对数据库的读取，Echarts对数据库中的回答量和阅读量读取，词云是对数据库中的问题或话题读取，经对比词云图使用问题数据和话题数据的显示结果相差不大。

5 系统测试

5.1 测试系统的目的

系统测试是运行程序以发现错误的过程，而成功的测试是发现以前未被发现的错误的过程。测试系统的目的是为了评估系统的功能和性能，确保系统在实际使用中能够正常运行、满足用户需求，并且能够在预期的工作环境下稳定运行。

验证系统的各项功能是否按照设计和需求进行实现。例如，测试系统可以检查系统是否正确获取知乎的话题数据，是否能够进行数据处理和分析，是否能够生成正确的可视化图表，是否能够进行交互等。

验证系统的性能，包括响应时间、处理数据的速度、系统的稳定性等。

验证系统的用户界面设计是否符合预期和需求，是否易于使用和操作。

5.2 测试方法

对爬虫数据可视化系统进行测试。对系统的不同模块和功能进行集成测试，验证其在整体系统中的协同工作和相互调用是否正确。

对整个系统全面的测试，验证系统的功能和性能是否符合设计和需求。可以通过模拟实际使用场景、进行用户操作测试。

表5-1 爬虫数据可视化系统的测试用例1

测试内容	测试操作描述	操作步骤	期望结果	实际结果
写入数据库	在把数据保存到数据库时报以下错误 7 values for 6 columns	观察后发现，数据库在存储数字时不需要加引号，而在回答量较多的情况下知乎数据在第三位前有逗号，如果不把逗号替换数据库就会以为传递了7个数据	能正常写入数据库	与期望结果一致
写入数据库	在写入数据库是若有些链接没内容，只有一个链接，写入数据库的列表只有一个值，写入数据库的有好几项，因此会写入错误。	重写代码：只写入数据库一项数据，列表只取第一位。 else:sql = '''insert into zhihu(link,wenti,huati, guanzhu,liulan,huida) values(%s,NULL,NULL, NULL,NULL,NULL) ''' % data[0]	没内容链接能正常写入数据库	与期望结果一致
重新执行	如果爬取过程由于某些原因中断，由于表已经建立，重新执行会报错。	用异常机制try:... except: print('已经存在表')	重写执行程序正常	与期望结果一致
已删除链接	例如在运行到518403259时出现提示410错误，打开改网页显示改内容已删除。	这种情况属于没有考虑到，解决办法：在if和else中间加上elif语句，同时修改if语句 if cuowu != 404 and cuowu != 410: # 判断页面是否为404 elif cuowu == 410:	能判断网页	与期望结果一致
没有用户回答	由于未考虑冷门问题没有人回答的现象出现以下错误 5 values for 6 columns 打开网页显示没有用户回答。	增加情况判断：在数据库的if-else中间加上elif。 elif len(data) == 5: # 没有回答的情况	能判断网页	与期望结果一致
写入数据库	提示：near “很人渣”:ayntax error 猜测：文字添加数据库是需要加引号，与数据的问题中的引号干扰	使用replace把双引号替换成单引号解决问题wenti = re.findall(rewenti, str(popover))[0].replace(',' , '')	能正常写入数据	与期望结果一致
网页图片	在编写展示数据网页时出现只有文字没有图片的情况。	分析原因由于访问路径改变导致图片和css路径错误，解决办法：加个/指向当前站点根目录中。批量替换路径。	能显示图片	与期望结果一致

测试内容测试操作描述操作步骤期望结果实际结果

写入数据库在把数据保存到数据库时报以下错误

7 values for 6 columns 观察后发现，数据库在存储数字时不需要加引号，而在回答量较多的情况下知乎数据在第三位前有逗号，如果不把逗号替换数据库就会以为传递了7个数据能正常写入数据库与期望结果一致

写入数据库在写入数据库是若有些链接没内容，只有一个链接，写入数据库的列表只有一个值，写入数据库的有好几项，因此会写入错误。 重写代码：只写入数据库一项数据，列表只取第一位。

else:sql = '''insert into zhihu(link,wenti,huati, guanzhu,liulan,huida) values(%s,NULL,NULL, NULL,NULL,NULL) ''' % data[0] 没内容链接能正常写入数据库与期望结果一致

重新执行如果爬取过程由于某些原因中断，由于表已经建立，重新执行会报错。 用异常机制try:... except: print('已经存在表') 重写执行程序正常与期望结果一致

已删除链接例如在运行到518403259时出现提示410错误，打开改网页显示改内容已删除。 这种情况属于没有考虑到，解决办法：在if和else中间加上elif语句，同时修改if语句

if cuowu != 404 and cuowu != 410: # 判断页面是否为404 elif cuowu == 410: 能判断网页与期望结果一致

没有用户回答由于未考虑冷门问题没有人回答的现象出现以下错误 5 values for 6 columns

打开网页显示没有用户回答。 增加情况判断：在数据库的if-else中间加上elif。elif len(data) == 5: # 没有回答的情况能判断网页与期望结果一致

写入数据库提示：near “很人渣” :ayntax error

猜测：文字添加数据库是需要加引号，与数据的问题中的引号干扰使用replace把双引号替换成单引号解决问题went_i = re.findall(rewent_i, str(popover))[0].replace('“', '”') 能正常写入数据与期望结果一致

网页图片在编写展示数据网页时出现只有文字没有图片的情况。 分析原因由于访问路径改变导致图片和css路径错误，解决办法：加个/指向当前站点根目录中。批量替换路径。 能显示图片与期望结果一致

5.3 测试结果

对与爬虫运行测试，Flask运行测试，词云生成图片测试，和Echarts图表显示测试。由于编写过程就是在不断的报错中修改的，所以最后的测试未出现错误。最终经测试爬虫程序能正常运行，Flask框架正常运行，Echarts能正常显示，WordCloud词云制作程序能正常输出图片。各个模块功能正常，模块间链接正常，系统整体正常。

表5-2 爬虫数据可视化系统的测试用例2

序号	测试点	操作步骤	期望结果	实际结果
1	爬虫运行测试	运行爬虫程序	运行成功，不报错，显示提示信息。	与预期相符
2	Flask运行测试	运行Flask程序	运行成功，在浏览器打开不报错。	与预期相符
3	词云生成图片测试	运行词云程序	运行成功，无报错，生成词云图。	与预期相符
4	Echarts图表显示测试	打开图表网页查看	打开成功，正常显示图表。	与预期相符

序号测试点操作步骤期望结果实际结果

- 1 爬虫运行测试运行爬虫程序运行成功，不报错，显示提示信息。 与预期相符
- 2 Flask运行测试运行Flask程序运行成功，在浏览器打开不报错。 与预期相符
- 3 词云生成图片测试运行词云程序运行成功，无报错，生成词云图。 与预期相符
- 4 Echarts图表显示测试打开图表网页查看打开成功，正常显示图表。 与预期相符

结论

本文设计并实现了Python网络爬虫，本文从爬虫和数据处理分析两部分来展开，完成了对各指定知乎话题精华问题的爬取，提取了问题的题目、话题、被浏览量、回答量等数据。使用此数据做出了观察浏览量与回答数量的散点图，并根据数据的话题做出了词云图。从知乎网站爬取了大量的数据。使用BeautifulSoup、正则表达、SQL处理、保存数据。基于多种第三方工具实现数据的可视化处理，具有很好的实用价值，通过分析数据提升了数据敏感度，提高了多维度看待问题的能力。

通过本设计能看出数据分析是从样本到总体意义上的推断、是精简过的，通过对原始数据的简单而直接的提取，虽然看不出这些迹象，但这也只是部分，而不是整体。数据分析的某些部分，超出了它的语言学范围，在某种意义上，他们指导我们观察或分析有价值的方向。数据分析是一个比推理的过程更大的环节。

后续工作的展望：由于时间关系爬取话题数量较少，后续爬取所有话题数据量高后要把这些连接起来做一个完全自动化的爬虫和数据展示分析，还有要做数据展示的排序，界面的美化，图表还要增加一个话题与关注者或者回答量的关系更精准的反应话题热度。

参考文献

[1]. 中国经济网. 2021新知青年大会开幕知乎将继续加大对创作者支持[J].

[2]. 潘娜. 基于大数据技术的电信客户维系挽留的分析与研究[D]. 河南:郑州大学,2017.

[3]. 张园园. 医疗贴吧中广告的提取系统[D]. 2016. DOI:10.7666/d.D01052856.

[4]. 孙立伟,何国辉,吴礼发. 网络爬虫技术的研究[J]. 电脑知识与技术, 2010, 6(15): 4112—4115.

[5]. 余洋. 豆瓣电影评论文本的情感分析及主题提取研究[D]. 云南:云南财经大学,2018.

[6]. 梁融凌,念其锋. 基于Python爬虫的胡润百富榜数据可视化分析[J]. 电脑与信息技术,2022, 30(06):46-50. DOI:10.19414/j.cnki.1005-1228.2022.06.031.

[7]. Yu L, Li Y, Zeng Q, et al. Summary of web crawler technology research[C]//Journal of Physics: Conference Series. IOP Publishing, 2020, 1449(1): 012036.

[8]. 陶皖主编. 云计算与大数据[M]. 西安电子科技大学出版社, 2017. 01: 第44页.

[9]. 杨磊磊. 大数据视角下非结构化文本数据的顾客满意度研究[D]. 2017. DOI:10.7666/d.D01198010.

[10]. Y. Daniel Liang. Python 语言程序设计[M]. 成都:机械工业出版社, 2013: 30-33.

[11]. Jennifer Campbell. 利用 python 进行数据分析[M]. 成都:机械工业出版社, 2012: 18-21.

[12]. 封俊. 基于 Hadoop 的分布式搜索引擎研究与实现[D]. [硕士学位论文]. 太原:太原理工大学, 2010

[13]. 崔蓬 .ECharts 在数据可视化中的应用 [J]. 软件工程 ,2019 (6) :42-46.

[14]. 张成强,卢凤娟. 项目管理平台的设计与实现[J]. 电脑知识与技术, 2021, 17(15):88-90. DOI:10.14004/j.cnki.ckt.2021.1466.

[15]. 杨健,陈伟. 基于Python的三种网络爬虫技术研究[J]. 软件工程, 2023, 26(02):24-27+19. DOI:10.19644/j.cnki.issn2096-1472.2023.002.005.

[16]. 曹阳. 我国对违反“爬虫协议”行为的法律规制研究[J]. 江苏社会科学, 2019(03):159-167. DOI:10.13858/j.cnki.cn32-1312/c.2019.03.021.

[17]. 阮泽楠. 音乐社交平台用户情绪特征研究[D]. 浙江理工大学, 2019.

[18]. 张瑾. 知乎“抄袭”话题评论的情感分析[D]. 云南财经大学, 2018.

致谢

四年的求学生涯，我走得虽然有点辛苦但是收获颇丰。在老师，朋友的全力支持下，也在不断地学习以及进步，在此论文

即将付梓之际，我也在反思自身，在这大学四年的光阴里，是否有虚度，是否有努力，是否有进步。

我的设计能够顺利进行离不开我的老师对我的指导。感谢IT私塾和Python爬虫课程教学老师，感谢你们免费的课程真正让人们平等接受教育，我学习到了更多的技能和学习方法。

同时，也非常感谢在我困难时给予我帮助与陪伴的伙伴们，你们的帮助与鼓励成为我坚持下来的力量。对此，最应该感谢的是在我身后默默支持我的父母，感谢我的家人在我学习生涯中给予我的无私帮助。感谢我的导师以及我的朋友，在我开始进入课题到最终论文得以完成对我给予了莫大关注，我真诚的感谢！

-
- 说明：1. 总文字复制比：被检测论文总重合字数在总字数中所占的比例
2. 去除引用文献复制比：去除系统识别为引用的文献后，计算出来的重合字数在总字数中所占的比例
3. 去除本人文献复制比：去除作者本人文献后，计算出来的重合字数在总字数中所占的比例
4. 单篇最大文字复制比：被检测文献与所有相似文献比对后，重合字数占总字数的比例最大的那一篇文献的文字复制比
5. 复制比：按照“四舍五入”规则，保留1位小数
6. 指标是由系统根据《学术论文不端行为的界定标准》自动生成的
7. 红色文字表示文字复制部分；绿色文字表示引用部分（包括系统自动识别为引用的部分）；棕灰色文字表示系统依据作者姓名识别的本人其他文献部分
8. 本报告单仅对您所选择的比对时间范围、资源范围内的检测结果负责



 amlc@cnki.net

 <https://check.cnki.net/>