

# 基于 Python 的高校网络舆情管理平台设计与实现研究

谢济江<sup>1</sup> 吴昊<sup>2\*</sup>

1 广西电力职业技术学院 2 南宁学院

DOI:10.32629/mef.v8i18.16987

**[摘要]** 本研究针对高校网络舆情监管需求,设计并实现了基于Python技术栈的舆情管理平台。该平台采用分层架构模式,集成Scrapy爬虫框架实现多源数据采集,运用自然语言处理技术进行情感倾向分析,构建智能预警机制实现舆情风险管控。系统以Django为核心框架,结合MySQL数据库和Redis缓存技术,实现了数据采集、情感分析、趋势预测、可视化展示等核心功能。通过实际部署测试,平台在数据处理准确率、响应速度和预警及时性方面表现良好,为高校舆情管理提供了有效的技术支持。

**[关键词]** 网络舆情; Python; 情感分析; 数据挖掘; 预警系统

中图分类号: V271.4+7 文献标识码: A

## Research on Design and Implementation of a Python-Based Platform for University Online Public Opinion Management

Jijiang Xie<sup>1</sup> Hao Wu<sup>2\*</sup>

1 Guangxi Electrical Polytechnic Institute 2 NanNing University

**[Abstract]** To address the public opinion monitoring needs of higher education institutions, this study developed and implemented a public opinion management platform based on Python technology stack. The platform adopts a layered architecture, integrating the Scrapy web crawling framework for multi-source data collection, employing natural language processing technology for sentiment analysis, and establishing an intelligent early warning mechanism to manage public opinion risks. Built on Django as the core framework with MySQL database and Redis caching technology, the system delivers core functionalities including data collection, sentiment analysis, trend prediction, and visual presentation. Through practical deployment and testing, the platform demonstrated excellent performance in data processing accuracy, response speed, and early warning timeliness, providing effective technical support for public opinion management in universities.

**[Key words]** online public opinion; Python; sentiment analysis; data mining; early warning system

### 引言

随着互联网技术的快速发展和社交媒体的广泛应用,网络舆情已成为影响高校稳定发展的重要因素。传统的人工监测方式已无法满足海量信息处理需求,迫切需要构建智能化的舆情管理体系。当前舆情监测系统存在数据源单一、分析深度不足、预警机制滞后等问题,难以准确把握舆情动态和发展趋势。基于此现状,本研究充分利用Python语言在数据处理和机器学习方面的优势,设计开发了集数据采集、智能分析、预警推送于一体的高校网络舆情管理平台,旨在提升舆情监测的自动化水平和分析精度,为高校管理部门提供科学决策依据。

### 1 高校网络舆情管理平台架构设计

#### 1.1 平台总体架构设计

本平台采用四层分布式架构模式,自底向上分别为数据采集层、数据处理层、业务逻辑层和用户展示层。数据采集层负

责从微博、贴吧、新闻网站等多个信息源获取原始数据,通过多线程爬虫和定时调度器实现24小时不间断数据收集,支持增量采集和全量更新两种模式,确保数据的实时性和完整性。数据处理层完成文本清洗、分词和特征提取等预处理工作,集成了去重算法、噪声过滤和数据标准化功能,同时建立了数据质量评估体系,对采集数据进行质量打分和异常检测。业务逻辑层实现情感分析、热点识别和趋势预测等核心算法,采用深度学习模型进行文本语义理解,结合统计学方法实现舆情态势分析,并通过机器学习算法构建预测模型。用户展示层提供可视化界面和交互功能,支持多终端访问和个性化定制,实现数据的直观展示和便捷操作。整体架构遵循松耦合设计原则,各层之间通过标准化接口进行数据交互,确保系统的扩展性和维护性。同时引入微服务理念,将舆情监测、数据分析、预警通知等功能模块化部署,采用容器化技术实现快速部署和弹性伸缩,提高系统的并发处理能

力和稳定性。此外，平台还集成了监控告警、日志审计、权限管理等辅助功能，为系统的安全稳定运行提供全方位保障。

### 1.2 核心功能模块设计

平台核心功能围绕舆情监测全流程设计，包含数据采集、智能分析、预警响应三大模块。数据采集模块基于Scrapy框架构建分布式爬虫系统，支持多线程并发采集和增量更新机制，日均处理数据量可达10万条以上，通过智能调度算法优化采集效率，避免重复抓取和资源浪费。该模块还集成了内容识别引擎，能够自动识别文本、图片、视频等多媒体内容，并对图片进行OCR文字识别，扩大监测范围。智能分析模块集成jieba分词工具和LSTM神经网络模型，实现中文文本的情感倾向判断，准确率达到85%以上，同时支持情感强度量化分析，能够识别轻微、中等、强烈等不同程度的情感表达。该模块还包含关键词提取、主题聚类、影响力评估等功能，通过多维度分析为舆情研判提供全面依据。预警响应模块建立多维度评估体系，设置三级预警阈值，当负面情绪占比超过60%或关注度增长率超过200%时触发自动预警，预警规则支持动态调整和个性化配置。各模块之间采用消息队列机制实现异步通信，通过Redis发布订阅模式实现实时数据流转，确保系统在高并发场景下的稳定运行<sup>[1]</sup>。此外，系统还设计了模块间的容错机制和降级策略，当某个模块出现异常时能够自动切换到备用方案，保证核心功能的持续可用。

### 1.3 技术选型与数据库设计

技术架构以Python生态系统为核心，充分利用其在数据科学和机器学习领域的优势，后端采用Django框架提供RESTful API服务，集成Django REST Framework实现接口标准化和权限控制，前端使用Bootstrap响应式框架和ECharts可视化库实现跨设备兼容的界面和丰富的数据展示效果。数据存储采用混合架构模式，MySQL主数据库存储用户信息、系统配置等结构化数据，Redis缓存热点数据和会话信息提升查询效率，Elasticsearch构建全文搜索引擎支持海量文本的快速检索和聚合分析，MongoDB存储非结构化的原始舆情数据满足灵活存储需求。数据库设计严格遵循第三范式，建立了舆情数据表、用户信息表、预警记录表、统计分析表等十余个核心表结构，通过外键约束和触发器保证数据一致性和完整性，设计了详细的索引策略优化查询性能。考虑到舆情数据的时效性特点和存储成本控制，设计了按时间维度的数据分区策略和定期归档机制，热数据保留3个月用于实时分析，温数据保留1年用于趋势分析，冷数据压缩存储用于历史回溯，有效平衡了查询性能和存储成本。整体技术选型充分考虑了开发效率、运行性能、维护成本和技术生态成熟度，为平台的长期稳定运行和功能扩展奠定了坚实基础。

## 2 舆情数据采集与智能分析实现

### 2.1 网络数据采集实现

网络数据采集系统基于Scrapy框架开发，采用分布式部署架构实现多源并发采集，支持横向扩展以应对数据量激增的需求。系统针对不同数据源制定专门的采集策略，建立了包含微

博、贴吧、论坛、新闻网站在内的20余个数据源适配器，每个适配器都针对目标网站的特点进行了深度优化。通过User-Agent池和代理IP轮换机制有效规避反爬虫检测，维护了包含200个不同浏览器标识的UA池和100个高质量代理IP的代理池，保证采集任务的连续性和稳定性。数据采集流程包括URL种子管理、页面解析、内容提取和数据清洗四个环节，种子管理模块通过广度优先搜索算法发现新的内容页面，并建立URL去重机制避免重复采集。页面解析模块采用XPath和CSS选择器相结合的方式提取目标信息，结合正则表达式处理特殊格式内容，确保数据完整性和准确性。为应对网站结构变化，系统设计了自适应解析机制和异常处理策略，通过机器学习算法自动识别页面结构变化，当单个数据源出现问题时能够自动切换备用方案或通知管理员手动调整。数据清洗模块通过正则表达式去除HTML标签和特殊字符，利用SimHash算法计算文本相似度去除重复内容，并对采集到的数据进行格式标准化处理，包括时间格式统一、编码转换、内容长度过滤等操作。系统还建立了采集效率监控机制，实时统计各数据源的采集成功率、响应时间和数据质量，为后续优化提供数据支撑，为后续分析提供高质量的数据基础<sup>[2]</sup>。

### 2.2 自然语言处理与情感分析

自然语言处理模块采用深度学习技术实现中文文本的情感倾向分析。文本预处理阶段使用jieba分词工具进行中文分词，结合自定义词典和停用词表提高分词准确性，随后通过词性标注和命名实体识别提取文本特征。情感分析模型基于双向LSTM神经网络构建，训练数据集包含10万条人工标注的情感样本，涵盖正面、负面、中性三种情感类别。模型训练过程采用词向量嵌入和注意力机制，有效捕捉文本的语义信息和情感特征。为提升模型的泛化能力，引入数据增强技术和集成学习方法，最终模型在测试集上的准确率达到87.3%。此外，系统还集成了关键词提取和主题挖掘功能，基于TF-IDF算法和LDA主题模型识别热点话题和关注焦点。

### 2.3 舆情趋势分析与预测

舆情趋势分析模块通过时间序列分析和机器学习算法实现舆情发展态势的预测。系统建立了包含情感指数、关注度、传播速度等多维度指标体系，采用滑动窗口机制计算各指标的动态变化趋势。时间序列分析采用ARIMA模型识别舆情数据的季节性和周期性规律，为短期预测提供依据。长期趋势预测则基于支持向量回归模型实现，通过历史数据训练得到预测模型，能够提前3-7天预警潜在的舆情风险。传播路径分析模块构建了信息传播网络图，运用社交网络分析方法识别关键传播节点和影响力用户，为舆情干预提供精准定位。预测结果通过可视化图表展示，包括趋势曲线、热力图和网络拓扑图等多种形式，帮助管理人员直观了解舆情演变规律和传播特征。

## 3 平台功能实现与应用效果评估

### 3.1 Web平台开发与部署

Web平台采用前后端分离架构，前端基于Vue.js框架开发单页面应用，集成Element-UI组件库实现界面的统一性和美观性，

通过ECharts图表库提供丰富的数据可视化效果。后端基于DjangoRESTFramework构建API服务,实现用户认证、权限管理、数据查询等核心功能,所有接口遵循RESTful设计规范确保系统的可扩展性。数据库连接池采用DjangoORM进行管理,通过索引优化和查询缓存提升数据访问效率。系统部署采用Docker容器化技术,结合Nginx反向代理和Gunicorn应用服务器实现负载均衡和高可用性。生产环境配置了Redis集群用于会话管理和数据缓存,Celery分布式任务队列处理异步作业,整体架构支持水平扩展以应对用户量增长需求<sup>[3]</sup>。

### 3.2 实时监控与预警系统

实时监控系统基于WebSocket技术实现数据的实时推送,用户可以在web界面上实时查看最新的舆情动态和分析结果。预警机制建立了三级风险评估模型,当系统检测到负面情绪激增、敏感词频繁出现或异常传播模式时自动触发预警流程。预警阈值设定采用动态调整策略,根据历史数据和季节性特征自动优化预警参数,减少误报和漏报现象。通知推送支持多种方式,包括邮件、短信和微信推送,管理员可以根据预警级别选择不同的通知渠道和响应策略。舆情报告生成模块提供自动化报告功能,系统每日定时生成舆情摘要报告,包含热点话题排行、情感趋势分析、重点事件跟踪等内容,支持PDF格式导出和邮件自动发送功能。

### 3.3 系统测试与应用效果分析

表1 系统性能测试结果测试指标统计表

测试项目	测试结果	性能基线	提升幅度
响应时间	2.3秒	4.1秒	43.90%
并发处理能力	1000用户	600用户	66.70%
数据处理速度	5万条/分钟	3万条/分钟	66.70%
预警准确率	80%	65%	23.10%
系统稳定性	99.20%	95%	4.40%

系统测试包含功能测试、性能测试和用户验收测试三个阶段,功能测试验证了数据采集、情感分析、预警推送等核心模块的正确性和稳定性。性能测试结果显示,系统在1000并发用户访问条件下响应时间保持在3秒以内,数据库查询效率较传统方案提升40%以上。实际应用测试选择了某高校一学期的舆情数据进

行验证,系统成功识别出15起潜在舆情事件,预警准确率达到80%,平均提前预警时间为6小时。用户满意度调研显示,90%的管理人员认为平台显著提升了舆情监测效率,界面友好性和功能实用性得到广泛认可<sup>[4]</sup>。基于用户反馈和实际应用经验,系统在算法精度、响应速度和用户体验方面持续优化改进,为高校网络舆情管理提供了有效的技术支撑。

## 4 结论

本研究成功设计并实现了基于Python的高校网络舆情管理平台,通过分层架构设计和模块化开发,构建了完整的舆情监测分析体系。平台集成了多源数据采集、智能情感分析、实时预警推送等核心功能,采用深度学习算法提升了情感识别准确率,通过分布式部署保证了系统的稳定性和扩展性。实际应用测试表明,系统在数据处理效率、预警及时性和用户体验方面均达到预期目标,为高校舆情管理工作提供了有力的技术支撑。然而,系统在处理网络语言和多模态数据方面仍有改进空间,未来可进一步优化算法模型,扩展数据源类型,提升舆情分析的深度和广度。研究成果对推动高校信息化建设和网络舆情治理现代化具有重要的理论价值和实践意义。

### [基金项目]

2022年度广西高校中青年教師科研基础能力提升项目,“基于Python的高校网络舆情管理平台开发”(项目编号:2022KY1333)。

### [参考文献]

[1]丁然.基于Python爬虫技术的高校网络舆情数据分析研究——以“安徽审计职业学院百度贴吧”为例[J].现代信息技术,2023,7(5):106-108.

[2]饶裕.基于python技术构建高校网络舆情预测模型应用[J].葡萄酒,2023(15):0133-0135.

[3]李文骞.基于Python的微博网络舆情监控系统设计与实现[J].电脑编程技巧与维护,2021,(012):49-51.

[4]彭健,周扬帆.基于Python的抖音“高职教育教学”舆情数据抓取与分析[J].科技资讯,2024,22(15):253-256.

### 作者简介:

谢济江(1986--),男,汉族,广西玉林人,本科,讲师,研究方向:人工智能。

### \*通讯作者:

吴昊(1991--),男,汉族,四川江油人,博士,讲师,研究方向:数据分析。