

# 毕业设计（论文）学生开题报告

课题名称	基于数据挖掘的舆情警报系统设计与实现				
课题类型	实践应用型			指导教师	王晓华
学生姓名	唐可寅	学号	201710253110	专业班级	17gb 计算 3 班

本课题的研究现状、研究目的及意义

1. 本课题的研究目的与意义

随着当代互联网和信息技术的飞速发展，网络技术的日新月异，网络已经逐渐融入到社会政治、经济生产和人民生活的各个方面，成为社会活动中不可或缺的一部分。信息技术的发展使得网络社交日益流行，并受到国内外广大群众的广泛关注和参与。每年通过在线社交网络注册的用户数都在不断增长。中国互联网络信息中心（CNNIC）于 2020 年 9 月 29 日发布第 46 次《中国互联网络发展状况统计报告》。报告显示，截至 2020 年 6 月底，我国总体网民规模达到 9.60 亿，互联网普及率为 67.0%。网络化生存已成为这个时代的基本能力和基本事实。

随着第四代、第五代的移动信息系统（4G、5G）的快速发展和应用，网络已经成为继报纸、广播、电视后的“第四媒体”，发挥着强大、高效的舆论表达和传递能力。社交网络为每一位网络参与者都提供了信息交互的平台，每一位网络参与者都有能力在网络上成为一名自媒体人，网络媒体可以成为一个自媒体平台，但他们不需要通过政府主管部门的审批，没有记者证、出版物刊号、运营许可证。当代中国正处于中华民族伟大复兴的关键时期，各种社会矛盾仍然比较突出、严重。网络作为一个传递信息的高效率平台，会使舆论信息快速传递并扩散，并形成舆情，对社会稳定带来重大影响。网络参与者逐渐深度介入各类社会事件的发展和演变过程，网络舆情已成为当代社会稳定和谐的一种标志，是各国政府和社会关注的焦点。近年来，社会事件频繁通过网络舆情进一步扩散发酵，成为社会矛盾爆发的导火索，对社会稳定带来巨大影响，使得网络热点事件分析与研究对网络舆情监控管理以及预警具有十分重要的现实意义。

本次毕业设计（论文）希望设计一种基于数据挖掘的舆情警报系统，使用计算机科学与技术行业各类前沿技术，并以大数据观念变革传统网络舆情管理思维，准确把握网络舆情的内在特征及其在演化过程中的潜在规律，对于新形势下做好网络舆情管理工作具有重要的理论意义和实践价值。

## 2. 本课题的研究现状：

舆论、舆情和网络舆情是一个具有中国特色的概念，英语语言体系中对舆论、民意、民情与舆情等都不做严格区分，一般都使用“Public Opinion”表述。西方国家民众观点的表达是政治生活中自然构成的一部分，渠道畅通，西方舆情研究起源于社会学领域，是传播学的研究内容之一，主要集中在民众观点的传播与调查的机理方面，如总统选举、国会选举等选情预测，顾客对企业产品质量口碑和态度调查，较少出现中国热点事件网络舆情汹涌现象。

随着互联网和自媒体的普及，越来越多的政府，企业开始意识到舆情监测的重要性，中国的舆情行业开始蓬勃发展。网络时代的突发重大事件中，“权威发布”难以抢到第一时间，表面上看是事发单位部门和相关媒体的问题，实际上是传统的新闻传播规律已经不适应自媒体时代的信息传播速度。

政府舆情相对发展较快，企业舆情多与公关或客服绑定在一起，因为一个政府要处理的舆情事件比一个企业要多得多，对舆情各方面的需求度都比较高。但不论政府还是企业舆情，基本都是由于内部前期事件没有处理好，才逐渐走向网络（排除突发公共事件，如天津大爆炸、6·1 东方之星旅游客船倾覆事件）。企业爆发的舆情，比如携程幼儿园事件、滴滴顺风车事件，都是内部发现问题处理，相关双方未达成和解，最终受害方寻求网络期望得到解决。也正是因此，企业舆情多侧重前期预防，而不是后期应对。而政府面对的突发舆情和不确定性，比企业相对要多。很明显会引发舆情的社会公共领域目前基本都已经显现，各个政府部门对此都有基本的认知，政府舆情目前更多是在后期应对，而不是前期预防。但是应对舆情的态度和处理水平有很大差别，从上到下，水平依次下降。

在线社会网络舆情是通过在线社会网络对各种社会现象、问题所表达的信念、态度、意见和情绪，往往具有极强的思想性和政治性，网络舆情一旦成为社会事件的放大器，凝聚共识，激发行动，必将影响社会。因此，无论是中央还是地方政府都非常重视网络舆情，科研院所和高校的学者从新闻学与传播学、图书馆情报与文献学、社会学、政治学、计算机以及这些学科的交叉等不同学科对网络舆情进行了相关研究。

关于网络舆情的分析与预警的研究，国外研究学者的研究方向主要集中在以下四方面：

- a. 针对互联网对舆情或网络舆情的影响展开；
- b. 侧重于危机事件应对和决策的研究，大多是以群体决策理论框架和批判理

论角度出发；

c. 针对群体传播理论的网络传播行为研究；

d. 针对网络传播群体中人员社会背景、性别、角色等由此带来等影响的研究，关注重点大多是不平等和压迫现象；

e. 针对公众舆情对政府及政府对重大事件政治决策的评论及其影响。

国内研究学者的研究方向主要集中在网络舆情监控预警理论体系建立和完善、网络舆情预警等级指标体系建立和完善、网络舆情预警的政府应对策略研究这 3 个方面。

课题类型： A-理论探究型；B-实践应用型。

## 本课题的研究内容

随着大量的舆论充斥网络，在互联网高速传播的帮助下，舆论已经成为一股非常强大的力量。由于网络具有虚拟的特点，一些非理性评论、负面报道或者谣言在网络上的传播极易激发网民的负面情绪，甚至会直接影响到社会大众对我国党以及政府的信任程度。因而，进一步深入研究基于数据挖掘技术的舆情监控系统的设计与实现具有十分重要的现实意义。

本次毕业设计（论文）将首先对与本课题相关的关键技术，以及国内外相关研究作出介绍，然后阐述在实际工作中遇到的问题，并结合用户需求，提出有效的解决思路 and 方案，包括将可能使用的工具、方法等。利用诸如数据挖掘、文本情感分析、大数据技术、机器学习等行业前沿信息化技术手段实现一套集“新闻、论坛、博客、微博客等互联网信息实时采集与发现；信息传播渠道、跟踪与溯源；舆情监测分析与预警、发展预测与防范、引导与干预”于一体的高效、智能化业务工作系统。

## 本课题研究的实施方案、进度安排

### 实施方案：

本次毕业设计（论文）为设计基于数据挖掘的舆情警报系统，将使用 Java(JDK 1.8)语言进行开发,使用 Spring Boot 或 Spring Cloud 作为 J2EE 框架进行系统开发,使用 MyBatis 和 MyBatis-Plus 作为数据库 ORM 框架，使用 Apache Shiro 和 JWT 作为权限管理和安全认证框架，使用 WebCollector、WebMagic、Htmlunit、jsoup 等网络爬虫框架技术模拟浏览器采集国内外各大社交媒体平台用户、推文、评论等信息。系统使用上述各类技术并结合行业当前前沿算法、技术监测社交平台上的各类推文信息，检测舆论信息，并给出信息反馈，从而实现本次毕业设计（论文）要求等系统功能和能力。

本次系统设计预期划分 6 个功能模块，以帮助实现系统整体功能，系统各功能模块设计目标如下：

- a. 数据采集模块：利用网络爬虫技术，爬取国内外各大社交媒体平台用户、推文、评论等信息，对数据编码、字段识别、字段抽取、数据去重等处理行动，最终数据入库；
- b. 数据预处理模块：依据采集到的数据信息进行数据清洗、提取、降噪、分词等数据预处理行动；

c. 数据分析模块：对数据预处理模块得到的预处理数据，利用全文检索、文本情感识别、聚类等技术采取关键词提取、摘要提取、舆论信息辨别、推文和舆论信息特征分析、用户和推文影响力评级等数据分析行动；

d. 舆情预警和人工干预模块：利用行业前沿技术分别对推文和舆论信息特征结合数据分析行动产生的结果进行学习和分析，并最终产生舆论报警预警信息，同时提供人工对机器识别分析结果进行干预的能力；

e. 信息展示、报送和对外接口模块：提供新闻关键字、舆情信息、舆情报警预警等信息向平台用户展示的能力，提供相关信息主动向主管、监管部门主动报送的能力，提供对外公开接口允许第三方获取本系统服务信息的能力；

f. 系统基础运行和管理模块：提供后台管理能力，如系统参数管理、账户管理、用户登陆注册管理、权限管理、新闻关键字管理、舆论管理等。

进度安排：

2020 年 12 月 18 号：学生完成选题；

2020 年 12 月 19 号 - 2020 年 12 月 31 号：由指导老师给学生下达任务书；

2021 年 1 月 4 号 - 2021 年 1 月 17 号：收集资料，完成开题报告撰写、外文文献翻译工作；

2021 年 1 月 18 号 - 2021 年 1 月 31 号：针对项目完成项目立项说明书、需求分析说明书、系统设计说明书（包含总体设计和详细设计）；

2021 年 2 月 1 号 - 2021 年 3 月 28 号（第 5 周前）：完成项目代码编写及完成对应的编码进度表；

2021 年 3 月 29 号 - 2021 年 4 月 25 号：完善项目、进行深入研究，撰写论文初稿

2021 年 4 月 25 号 - 2021 年 5 月 14 号：指导教师对初稿进行批改，检查参考文献阅读情况，学生根据指导意见对初稿进行修改，并对项目进行进一步的完善。

2021 年 5 月 15 号 - 2021 年 5 月 21 号：论文定稿，项目完成，答辩资格审查，答辩资料准备，完成毕业设计（论文）重复率自查工作，并提供查重报告。

已查阅的主要参考文献

## 1. 会议类

- [1] Ge Hongbing. Research on the Construction of Early Warning Index System of Network Public Opinion Emergency Based on Computer Simulation[A]. 安徽工业大学.Proceedings of 2019 2nd International Conference on Manufacturing Technology, Materials and Chemical Engineering (MTMCE 2019)[C].安徽工业大学:香港环球科研协会,2019:6.
- [2] Yongchang Ren,Zhe Li. Internet Public Opinion Management in Big Data Era[A]. Institute of Management Science and Industrial Engineering.Proceedings of 2018 5th International Conference on Education,Management,Arts,Economics and Social Science(ICEMAESS 2018)[C].Institute of Management Science and Industrial Engineering:计算机科学与电子技术国际学会(Computer Science and Electronic Technology International Society),2018:5.
- [3] Yong-qing YUAN,Qin-ying WANG,Feng-ming LIU. Analysis of Internet Public Opinion Situation[A]. Advanced Science and Industry Research Center.Proceedings of 2018 International Conference on Modeling, Simulation and Analysis(ICMSA 2018)[C].Advanced Science and Industry Research Center:Science and Engineering Research Center,2018:6.
- [4] Fulian Yin,Beibei Zhang,Ge Su,Ruizhe Zhang. Research on the Public Opinion Pre-warning Based on the Logistic Model[A]. The Institute of Electrical and Electronics Engineers、IEEE Beijing Section.Proceedings of 2017 IEEE 8th International Conference on Software Engineering and Service Science[C].The Institute of Electrical and Electronics Engineers、IEEE Beijing Section:IEEE BEIJING SECTION(跨国电气电子工程师学会北京分会),2017:5.
- [5] MA Jing. Analysis on Governance of Government Online Public Opinion Big Data Perspective[A]. University of Ghana (UG)、Ghana Institute of Management and Public Administration(GIMPA)、University of Cape Coast (UCC)、American Society for Public Administration (ASPA)、University of Electronic Science and Technology of China (UESTC).Proceedings of 2017 International Conference on Public Administration(12th) & International Symposium on West African Studies (1st)(Volume II)[C].University of Ghana (UG)、Ghana Institute of Management and Public Administration(GIMPA)、University of Cape Coast (UCC)、American Society for Public Administration (ASPA)、

University of Electronic Science and Technology of China (UESTC):公共管理国际会议组委会,2017:6.

- [6] Xiaohong Jin,Hui Li,Hong Fu,Jing Zhang,Boda Ding. A Study on Early-Warning Response of Emergency Events Network Public Opinion Based on Analysis of Microbloggers' Behaviors[A]. IEEE Beijing Section, China、上海财经大学、同济大学.Proceedings of the 2016 IEEE International Conference on Progress in Informatics and Computing(PIC)[C].IEEE Beijing Section, China、上海财经大学、同济大学:IEEE BEIJING SECTION(跨国电气电子工程师学会北京分会),2016:4.
- [7] Zhao Liyong,Zhang Xu,He Min,Zhang Dan,Liu Wei,Liu Chunyang. Research on Public Opinion Index System of Chinese Microblog[A]. IEEE Beijing Section.Proceedings of 2014 IEEE 5th International Conference on Software Engineering and Service Science[C].IEEE Beijing Section:IEEE BEIJING SECTION(跨国电气电子工程师学会北京分会),2014:4.

## 2. 期刊论文

- [1] 杨菁,赵岩.网络舆情及其预警研究综述[J].江苏科技信息,2013(06):12-13.
- [2] 杨娜,褚茂胜.基于深度学习的网络舆情情绪监测系统的设计[J].中国科技信息,2020(24):76-78.
- [3] 郑昌兴.关于突发事件网络舆情监测预警的探讨[J].电脑知识与技术,2020,16(33):66-68.
- [4] 梁泽鹏,李直达.基于机器学习的舆情预警模型的设计与实现[J].广东公安科技,2020,28(03):1-5.
- [5] 杨森,王黎,李超.基于主动监测引擎和大数据的网络舆情烈度演化分析系统[J].价值工程,2020,39(17):201-203.
- [6] 曾宇.基于大数据的网络舆情实时监测系统的构建[J].漳州职业技术学院学报,2020,22(02):92-99.
- [7] 段峰峰,陈淼.多元与智能:跨媒体网络舆情数据的分析与处理[J].传媒观察,2020(01):33-38.
- [8] 万倩,朱里越,欧阳峰.基于人工智能的广电舆情分析系统[J].广播与电视技术,2019,46(12):46-52.
- [9] 刘小满,王小辉.基于“互联网+”的网络舆情监控系统的设计与实现[J].电脑知识与技

术,2019,15(31):37-40.

- [10] 白茹.基于云计算和 Hadoop 的网络舆情监控系统设计[J].电子设计工程,2019,27(16):141-144+150.
- [11] 洪小娟,宗江燕,于建坤,黄卫东.网络舆情监测系统的分析与设计[J].软件工程,2019,22(08):37-39+13.
- [12] 祁凯,彭程.基于 OCS-EGM 模型的网络集群行为监测及预警体系研究[J].情报杂志,2019,38(09):134-141+149.
- [13] 江瑾.网络舆情监控系统的设计和实现[J].信息与电脑(理论版),2019(13):63-65.
- [14] 王晨妮,王宇晨,张超,刘蓝静,孙钜晖,宋威.基于非结构化数据处理的网络舆情监测系统[J].科技创新与应用,2019(13):38-40.
- [15] 李猛哲,杨跃东.教育考试舆情监测指标体系设计[J].中国考试,2019(04):35-40.
- [16] Jian Li. Analyse the data tendency in the public opinion monitoring system[J]. Journal of Physics: Conference Series,2019,1187(5).
- [17] 潘晓英,陈柳,余慧敏,赵逸喆,肖康泞.主题爬虫技术研究综述[J].计算机应用研究,2020,37(04):961-965+972.
- [18] Chuiju You,Dongjie Zhu,2,Yundong Sun,Anshan Ye,Gangshan Wu,Ning Cao,Jinming Qiu,Helen Min Zhou. SNES: Social-Network-Oriented Public Opinion Monitoring Platform Based on ElasticSearch[J]. CMC: Computers, Materials & Continua,2019,61(3).
- [19] 陈刚,李弼程,郭志刚,林琛.网络舆情监测预警系统模型与关键技术[J].信息工程大学学报,2019,20(01):116-121.
- [20] 董积有,李岸,王海艺.基于深度学习的高校舆情监测的探索与研究[J].计算机产品与流通,2017(08):195.
- [21] 张荣显,曹文鸳.网络舆情研究新路径:大数据技术辅助网络内容挖掘与分析[J].汕头大学学报(人文社会科学版),2016,32(08):111-121.
- [22] 刘丽员,杨昔阳.基于文本相关性的高校网络舆情监控系统的设计与实现[J].泉州师范学院学报,2016,34(02):50-54.
- [23] 柳向东,曹雨婷,李利梅.网络影响力预知模型:一种大数据下高校舆情监测与预警机制[J].深圳大学学报(人文社会科学版),2015,32(04):156-160.



- [24] 于娟,刘强.主题网络爬虫研究综述[J].计算机工程与科学,2015,37(02):231-237.
- [25] 方星星,鲁磊纪,徐洋.网络舆情监控系统中主题网络爬虫的研究与实现[J].舰船电子工程,2014,34(09):104-107.
- [26] 刘红梅.垂直搜索引擎主题爬虫搜索策略研究[J].科技信息,2013(24):252-253.
- [27] 石鲁生,陈林,李凯.一种网络舆情的动态预警方法[J].天津师范大学学报(自然科学版),2012,32(02):59-65.
- [28] 张瑞.SMS 网络舆情信息监控系统的设计与实现[J].现代情报,2012,32(03):68-71.
- [29] 王长宁,陈维勤,许浩.对微博舆情热度监测及预警的指标体系的研究[J].计算机与现代化,2013(01):126-129.
- [30] 宋海洋,刘晓然,钱海俊.一种新的主题网络爬虫爬行策略[J].计算机应用与软件,2011,28(11):264-267+293.
- [31] 叶昭晖,曾琼,李强.基于搜索引擎的网络舆情监控系统设计与实现[J].广西大学学报(自然科学版),2011,36(S1):302-307+334.
- [32] 高承实,荣星,陈越.微博舆情监测指标体系研究[J].情报杂志,2011,30(09):66-70.
- [33] 魏晶晶,杨定达,廖祥文.基于网页内容相似度改进算法的主题网络爬虫[J].计算机与现代化,2011(09):1-4.
- [34] 张立涛,綦振法.“五位一体”的网络舆论监控体系研究[J].中国软科学,2010(07):8-15.
- [35] 唐铁兵,陈林,祝伟华.基于 Lucene 的全文检索构件的研究与实现[J].计算机应用与软件,2010,27(02):197-199+230.
- [36] 施聪莺,徐朝军,杨晓江.TFIDF 算法研究综述[J].计算机应用,2009,29(S1):167-170+180.
- [37] 陈丛丛. 主题爬虫搜索策略研究[D].山东大学,2009.
- [38] 黄旭,朱艳琴,罗喜召.基于内容评价的爬虫搜索策略研究[J].微电子学与计算机,2008(11):25-28.
- [39] 倪贤贵,蔡明.基于链接结构和内容相似度的聚焦爬虫系统[J].计算机工程与设计,2008(07):1709-1710+1763.
- [40] 蒋宗礼,徐学可,李帅.一种基于超链接引导的主题搜索的主题敏感爬行方法[J].计算机应用,2008(04):942-944+950.
- [41] 王娟.网络舆情监控分析系统构建[J].长春理工大学学报(高教版),2007(04):201-203+214.

- [42] 毛欣娟.网络舆论控制机制的构建及相关问题思考[J].江西公安专科学校学报,2007(06):49-52.
- [43] 刘金红,陆余良.主题网络爬虫研究综述[J].计算机应用研究,2007(10):26-29+47.
- [44] 梅中玲.基于 Web 信息挖掘的网络舆情分析技术[J].中国人民公安大学学报(自然科学版),2007(04):85-88.
- [45] 初建崇,刘培玉,王卫玲.Web 文档中词语权重计算方法的改进[J].计算机工程与应用,2007(19):192-194+198.
- [46] 曾凡斌.重大突发事件中的 BBS 舆论特点与管理初探——对人民网“强国论坛”的个案观察[J].出版发行研究,2006(04):61-67.
- [47] 王甘武.用辩证的方法正确把握网上舆论导向[J].信息网络安全,2006(04):28-30.
- [48] 李正华.网络舆论监督的临界问题及相关法律责任[J].信息网络安全,2006(04):31-33.
- [49] 叶惠敏,戴冠中.基于综合集成方法的网上舆论倾向分析与评估系统方案[J].计算机工程与应用,2005(16):216-217+220.
- [50] 叶惠敏,唐三平.用于网上舆论观点抽取的几种方法[J].计算机应用研究,2005(05):256-257+260.
- [51] 张校乾,金玉玲,侯丽波.一种基于 Lucene 检索引擎的全文数据库的研究与实现[J].现代图书情报技术,2005(02):40-43+48.
- [52] 李晓明,朱家稷,闫宏飞.互联网上主题信息的一种收集与处理模型及其应用[J].计算机研究与发展,2003(12):1667-1671.
- [53] 鲁松,李晓黎,白硕,王实.文档中词语权重计算方法的改进[J].中文信息学报,2000(06):8-13+20.

### 3. 硕博论文

- [1] 王鑫. 基于 BS 模式的网络舆情监控系统设计与实现[D].电子科技大学,2020.
- [2] 段吉东. 基于热量模型的微博舆情实时监控系统设计研究[D].济南大学,2020.
- [3] 胡中皓. 基于规则和情感的法治舆情监测系统研究与实现[D].江西财经大学,2020.
- [4] 钱晨. 基于信息测度的舆情情感态势模型研究[D].南京邮电大学,2019.
- [5] 王晶. 关于舆情监测系统实时交互式可视化的研究[D].西安理工大学,2019.
- [6] 赵康. 面向主题的网络爬虫系统的设计与实现[D].北京邮电大学,2019.

- [7] 李寒冰. 基于深度学习的舆情监测系统[D].北京交通大学,2019.
- [8] 王天齐. 基于微博的涉警网络舆情风险评估模型研究[D].中国人民公安大学,2019.
- [9] 李迁. 公安部门互联网舆情监管系统设计与实现[D].大连海事大学,2017.
- [10] 柳淑婷. 基于社交网络的舆情关键技术研究[D].吉林大学,2017.
- [11] 赵丽梅. 网络舆情分析与监测系统的设计与实现[D].兰州大学,2017.
- [12] 黄宝成. 基于社交网络的舆情信息挖掘方法研究[D].哈尔滨工业大学,2017.
- [13] 李洪胜. 内容与链接结构相融合的主题爬虫研究[D].华侨大学,2016.
- [14] 林红静. 基于 K-means 的微博短文本聚类算法研究[D].海南大学,2016.
- [15] 罗路天. 垂直搜索引擎中主题网络爬虫算法的设计与研究[D].广东工业大学,2016.
- [16] 邹妍. 网络舆情监控与分析系统的设计与实现[D].吉林大学,2015.
- [17] 冯佳明. 深度词汇网络学习舆情监测关键技术研究[D].北京化工大学,2015.
- [18] 邹建明. 面向广东基层公安的网络舆情监控系统设计与实现[D].电子科技大学,2014.
- [19] 张蕾. 基于机器学习的网络舆情采集技术研究与设计[D].电子科技大学,2014.
- [20] 董坚峰. 面向公共危机预警的网络舆情分析研究[D].武汉大学,2013.
- [21] 夏景隆. 基于版权服务的网络数据采集算法研究[D].北方工业大学,2013.
- [22] 周小丽. 基于网络爬虫和 Lucene 索引的互联网舆情监测系统设计与实现[D].吉林大学,2013.
- [23] 侯万友. 群体性突发事件微博舆情演化分析[D].哈尔滨工业大学,2013.
- [24] 吴健. 基于 Hadoop 的上市公司舆情挖掘系统的研究与实现[D].电子科技大学,2013.
- [25] 万源. 基于语义统计分析的网络舆情挖掘技术研究[D].武汉理工大学,2012.
- [26] 李斌. 基于 Lucene 的分布式全文检索在构件库系统中的实现[D].西安电子科技大学,2012.
- [27] 巨慧慧. 针对网络舆情的情感倾向性研究[D].哈尔滨工业大学,2010.
- [28] 蒋昌金. 基于关键词提取的中文网页自动文摘方法研究[D].华南理工大学,2010.
- [29] 杜阿宁. 互联网舆情信息挖掘方法研究[D].哈尔滨工业大学,2007.
- [30] 郑健珍. 定题爬虫搜索策略研究[D].厦门大学,2007.
- [31] 刘云峰. 基于潜在语义分析的中文概念检索研究[D].华中科技大学,2005.

指导教师意见

指导教师签名：  
年 月 日