**网络评论过滤与分析系统**

**研究报告**

[一 选题背景 3](#_Toc2188530)

[*（一）* *研究目的* 3](#_Toc2188531)

[*（二）研究背景* 3](#_Toc2188532)

[*（三）研究意义* 5](#_Toc2188533)

[二 、实施进程 5](#_Toc2188534)

[三 、成果内容 7](#_Toc2188535)

[*（一）项目内容简介* 7](#_Toc2188536)

[1.1热点话题、敏感话题识别 7](#_Toc2188537)

[1.2评论主题追踪 7](#_Toc2188538)

[1.3自动摘要 7](#_Toc2188539)

[1.4评论趋势分析 7](#_Toc2188540)

[1.5评论统计报告 7](#_Toc2188541)

[*（二） 项目实际内容* 7](#_Toc2188542)

[2.1网络评论过滤与分析系统web端 7](#_Toc2188543)

[2.2 PC端分析报告图 7](#_Toc2188544)

[四、创新点 8](#_Toc2188545)

[*（一）创新点* 8](#_Toc2188546)

[1.1技术创新 8](#_Toc2188547)

[1.2使用便捷安全 8](#_Toc2188548)

[1.3信息公开透明 8](#_Toc2188549)

[1.4 自动报告文档 8](#_Toc2188550)

[*（二）特色* 8](#_Toc2188551)

[2.1前沿性 8](#_Toc2188552)

[2.2 实用性 8](#_Toc2188553)

[2.3开放性 8](#_Toc2188554)

[2.4安全性 8](#_Toc2188555)

[*（三）学生主导表现* 8](#_Toc2188556)

[3.1分工明确 8](#_Toc2188557)

[3.2 独立开展 8](#_Toc2188558)

[3.3 积极讨论 8](#_Toc2188559)

[五、管理战略 9](#_Toc2188560)

[5.1项目概述 9](#_Toc2188561)

[5.2利益相关群体 9](#_Toc2188562)

[5.3 项目发展战略 9](#_Toc2188563)

[*六、市场营销* 10](#_Toc2188564)

[6.1目标市场 10](#_Toc2188565)

[6.2服务 10](#_Toc2188566)

[6.3 品牌 10](#_Toc2188567)

[6.4 经营渠道 10](#_Toc2188568)

[6.5推广策略 10](#_Toc2188569)

[七、 成果应用情况 10](#_Toc2188570)

[*（一）成果应用情况* 10](#_Toc2188571)

[*（二）应用前景* 11](#_Toc2188572)

[八、研究心得 11](#_Toc2188573)

[*（一）创新思维与成长的收获* 11](#_Toc2188574)

[*（二）可借鉴的经验* 11](#_Toc2188575)

# 一 选题背景

## *研究目的*

随着互联网的快速发展，网络媒体作为一种新的信息传播形式，已深入人们的日常生活。网友言论活跃已达到前所未有的程度，不论国内外重大事件，都能马上形成网上舆论。好的舆论可以成为发展的“推进器”，不好的舆论可以成为民众的“迷魂汤”、社会的“分离器”、杀人的“软刀子”、动乱的“催化剂”。由此可见，坚持正确的舆论导向，建设积极健康的网络空间必不可少。因此此次我们搭建网络评论过滤与分析系统，一方面期望净化评论环境，另一方面能够有效整理并分析评论，并加以正确应对和引导。

网络评论过滤与分析系统，通过对网页评论内容的采集处理、敏感词过滤、智能分类、主题检测、专题聚焦、统计分析，实现各团体及个人对自己相关网络舆情监督管理的需要，最终形成舆情简报、舆情专报、分析报告、移动快报，为决策层全面掌握舆情动态，做出正确舆论引导，提供分析依据。

## *（二）研究背景*

**1.社会现状分析：**

党的十八大以来，国家进入全面建成社会主义现代化强国的新征程。推动社会主义文化繁荣昌盛，牢牢掌握意识形态工作领导权具有重大意义。要掌握意识形态工作领导权，要建设好网络空间。互联网不是法外之地，要加强建设和管理，要本着对社会负责、对人民负责的态度，建立网络综合治理体系。尤其要加强网上舆论引导，培育向上向善的网络文化，营造一个风清气正的网络空间。

在当今新媒体强势兴起和传统媒体非常活跃的环境下，舆论场一方面呈现出自由开放的态势，有利于公民对公权力行为的监督，使一些腐败官员丢掉乌纱帽，受到了法律的制裁，媒体有效地行使了舆论监督的作用；另一方面不实信息混淆其中，甚至有的借机造谣，迷惑受众，误导社会舆论。舆论监督是新闻传播的基本功能，不仅代表着媒体的良知，也昭示着我国民主法治的进程。舆论对社会的导向性可见一斑.由此把握好芜杂信息的处理原则，去伪存真，净化舆论监督环境不可忽视。

可见，在当今的媒体时代，舆论监督的重要性不可忽视。因为它直接来源与群众，却又有区别于群众的呼声。同时我们应该注意到，近年来随着微博、贴吧、微信公众号等自媒体平台的不断发展，舆论导向不仅仅局限于新闻中，在这些允许用户自由讨论的、更贴近用户平常生活的平台中，面向用户的层次不同，评论内容也有着观点鲜明的特征，使得这些评论有着极大的研究价值。而如今大多数研究还局限在对新闻舆论的监督与分析，很容易忽略这些贴近生活，时下兴起的用户评论。此外，由于评论基数大，质量参差不齐，对这些评论不能很好地统计与分析。亦或是只能窥见其冰山一角，以偏概全。

开展这个项目之前我们对相关问题进行了一些调研，具体调研结果如下：

参加数据调研者150人，**分别**来自各行各业

**2.研究现状分析：**

国内外已有对评论文本的特点、文本的句法结构、机器学习等方法对垃圾内容的检测识别进行研究。Jindal等认为垃圾文本内容具有重复性，故通过识别评论中重复或近似重复的内容与评论对象的语义关系进行识别；刁宇峰等利用LDA模型对博客数据进行主题聚类，将博客文本包含的与垃圾评论相关的词汇聚类为垃圾主题，并用主题形式对隐含垃圾观点进行刻画，提高垃圾观点检测的准确率；Mukherjee等够早了正对在线评论的与LDA类似的产生式模型，加入了评论作者、文本、评分等生成关系，提高了垃圾主题聚类性能

## *（三）研究意义*

在这个背景下，网络评论过滤与分析系统呼之欲出。该系统通过对热点问题和重点领域比较集中的网站信息，如：网页、论坛、BBS等，随时爬取最新的消息和意见。爬取后完成对数据格式的转换及元数据的标引。借鉴已有的工作中对文本特征与垃圾评论相关关系规律的总结，结合基于规则的分类模型以及神经网络架构，对爬取至本地的信息，进行初步的过滤和预处理。对热点问题和重要领域实施监控，此外，对信息阐述的观点、主旨进行倾向性分析，以提供参考分析依据。分析的依据可根据信息的转载量、评论的回复信息时间密集度来判别信息的发展倾向

通过网络评论过滤与分析系统，相关部门可以及时了解网络舆情动态，关注到自己在网络舆情中的状态，从而可以产生网络舆情预警，及时纠正应对网络上关于自己的负面舆论影响，为部门网络危机公关或品牌形象营销提供数据依靠。对于个人用户与研究机构而言，该系统可以帮助用户高效、完整的了解网络上的其他用户对各种事物的评论数据统计。

# 二 、实施进程

2018年10-11月：

学习相关技术，HTML5、CSS、XML、JavaScript、PHP、SQL等，查阅相关开发书籍。

爬取大量数据，搜集大量数据集与论文，对评论筛选模型进行设计与测试。其中查阅NLP内相关知识，联系被试人员对一些无效评论进行人为设计以及分类，并测试筛选模型。

2018年12月至2019年1月：

筛选模型的进一步测试，以及研制评论分类模型。由上一阶段的成果先进行划分，并在一些主领域进行分类（如支持、反对等）。收集数据集对模型进行优化，同时考虑到人文因素，仍需联系被试人员进行人为测试以及准确率检验

2019年2月至2019年3月：

将得出的数据结果进行统计分析或者是采用决策分析算法，通过阅读有关统计知识采取科学的展示方式，数据库设计与建表。

2019年4月至2019年6月：

优化系统操作体验。考虑联系部分媒体以及商业公司。试采用该模型进行信效度检验，测试在现实情况下的使用情况以及是否存在疏漏等。

2019年7月至2019年8月：

对模型进行优化以及修复bug.进入测试阶段，主要通过邀请学生参与试用评测的方式，记录每次测试的相关数据，根据数据结果对出现的bug或不足之处进行修改加工，并拿出最终的成品。

2015年9月:

开始撰写关于本项目的研究报告，其中将包括整个研发过程的概述，最初的想法，初步的构思，明确框架，着手研究等，重点在于技术实现过程中遇到的难题及解决办法，最后有小组成员的心得体会等。进入结题及答辩阶段，向组织方提供完整的项目成果，并就专家组的问题进行答辩。

# 三 、成果内容

## *（一）项目内容简介*

### 1.1热点话题、敏感话题识别

### 1.2评论主题追踪

### 1.3自动摘要

### 1.4评论趋势分析

### 1.5评论统计报告

## *（二） 项目实际内容*

### 2.1网络评论过滤与分析系统web端

### 2.2 PC端分析报告图

# 四、创新点

## *（一）创新点*

### 1.1技术创新

### 1.2使用便捷安全

### 1.3信息公开透明

### 1.4 自动报告文档

## *（二）特色*

### 2.1前沿性

### 2.2 实用性

### 2.3开放性

### 2.4安全性

## *（三）学生主导表现*

### 3.1分工明确

### 3.2 独立开展

### 3.3 积极讨论

# 五、管理战略

### 5.1项目概述

#### 5.1.1 项目基本情况

#### 5.1.2 项目主要特征

#### 5.1.3 组织结构

### 5.2利益相关群体

#### 5.2.1 在校学生

#### 5.2.2 研究机构及组织、团体

#### 5.2.3 高校

#### 5.2.4 公关人员

#### 5.2.5 政府部门

### 5.3 项目发展战略

#### 5.3.1 项目初期（1-2年）：建设团队，树立品牌

#### 5.3.2项目中期（2-5年）：建立业务网，有效经营

#### 5.3.3项目成熟期（5-10年）：扩张网点，扩大服务对象

## *六、市场营销*

### 6.1目标市场

### 6.2服务

#### 6.2.1 对校内学生

#### 6.2.2 对历届校友校友

### 6.3 品牌

#### 6.3.1 以成功的项目提升品牌

#### 6.3.2 以完善的服务培养忠诚度

### 6.4 经营渠道

#### 6.4.1建立以华东师范大学为中心的经营网

#### 6.4.2利用互联网等新媒体平台

### 6.5推广策略

# 七、 成果应用情况

# （一）成果应用情况

网络评论过滤与分析系统即将正式上线web平台，将先在本校师生内预推广，吸取并采纳用户意见，再逐步推广至社会群体，以免费的形式提供给学生、团体、社会各界人士进行推广使用。

## *（二）应用前景*

项目网站目前搭建在“阿里云”服务器上，其安全性与效率得以保障，同时使用部署在服务器上的MySQL数据库，能够在保证用户体验的同时支持大量访问。

网络评论过滤与分析系统不仅能够满足一般用户的基本要求，对不同的用户又有着不用的适应性特点。其本身方便快捷、安全放心、功能完备。项目首页搜索页面简单易操作，支持多种网页格式并提供大量链接方便不同用户进行查询。后台管理页面能及时对最新舆论事件进行更新，查询并搜集评论数据，生成报告存入数据库，方便管理员整理以及用户需求。报告页面功能完备齐全，并具有一定可扩展性，大大节省了用户的时间和搜集查询的烦恼。此外其本身免费、方便等特征，更加吸引和迎合不同年龄段的用户。在校学生能够更好地分析和把握舆论动向，统计和分析评论用户的情感动向；诸如公关人员的社会人士能时刻把握评论动向，更好地应对和处理人员事件；政府部门等团体可以加强对网络评论的及时监测、有效引导，以及对网络舆论危机的积极化解，对维护社会稳定、促进国家发展具有重要的现实意义，也是创建和谐社会的应有内涵。

综上本项目并未使用成本很高的软硬件设备，其本身免费、开放的特性将不断地吸引更多用户参与使用并进行功能优化提升，放在实际使用中可以预计会有较高的实用价值与社会效益。

# 八、研究心得

## *（一）创新思维与成长的收获*

首先非常感谢学校为我们提供了一次自主创新，深入课题研究的机会。在这次团队合作、自我锻炼、在实践中强化编码能力的过程中，我们收获了很多，也有一些心得体会。

在学习方面，通过这次项目强化了我们学习新知识的能力，使我们更深入地了解网站结构的搭建，使我们积累更多对模型的搭建与运用的经验，使我们的知识结构更全面。通过从相关论文中学习他人宝贵经验以及在其基础上进行改进与创新，这对于我们培养创新性思维和严谨的研究态度大有裨益。并且，在每一次调试参数的过程中，都积攒了许多自主学习自主解决问题的经验。

在团队合作方面，我们在课题初期需求并不完全确定的时候比较迷茫，但在谢瑾奎教授的宏观指导和细节指点下，一次次会议让我们明确了需求和功能模块，也明确了努力的方向；同学之间不断交流彼此心得，互相帮助解决遇到的问题，不仅学习到很多知识，还让彼此间建立了深厚的友谊。这对于我们今后日常学习、团队合作乃至将来的日常工作都有着不可估量的启迪。

## *（二）可借鉴的经验*

在项目编码正式开始之前，首先最重要的是要把需求明确下来：包括功能模块的设计、前端页面的设计、模型的设计、数据库的设计，更细化的比如：代码包、类的方法和成员函数的设计等等，由于每位成员负责的板块有所不同，如果一开始这些问题比较模棱两可，并没有严格的规定，那么到了后期改动会比较麻烦，而且往往越底层的修改越麻烦。

编码过程中当自己原有的方法无法解决问题时，需要积极的学习新知识而不是拘泥于自己所熟悉的方法，只有这样才能推动项目的进展，同时也提高自身的技术水平。另外在编码过程中遇到困难时，除了从有关文献、资料中搜集现有的方法以外，及时与团队交流非常重要，有时候来自不同视角的看法可以提供新的解决方案，也可以避免网上一些良莠不齐的“答案”对我们学习的影响。

当项目编码内容基本完成后，测试以及部署是保证项目正常运行的极其重要的两个步骤。测试需要注意的问题是，集中测试后将需要修改的地方一起重新编码，并编以版本号，这样一来方便查看每个版本修改的问题。在部署的时候，要注意服务器的环境并做好记录，必要时需要做端口映射。