

学校编码: 10384

分类号_____密级_____

学号: X2012230912

UDC_____

廈門大學

工程硕士学位论文

网络舆情监测系统的设计与实现

Design and Implementation of Network Public Opinion
Monitoring System

吴愧

指导教师: 陈海山教授

专业名称: 软件工程

论文提交日期: 2015 年 月

论文答辩日期: 2015 年 月

学位授予日期: 年 月

指导教师: _____

答辩委员会主席: _____

2015 年 月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下，独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果，均在文中以适当方式明确标明，并符合法律规范和《厦门大学研究生学术活动规范（试行）》。

另外，该学位论文为（ ）课题（组）的研究成果，获得（ ）课题（组）经费或实验室的资助，在（ ）实验室完成。（请在以上括号内填写课题或课题组负责人或实验室名称，未有此项声明内容的，可以不作特别声明。）

声明人（签名）：

年 月 日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

1.经厦门大学保密委员会审查核定的保密学位论文，于
年 月 日解密，解密后适用上述授权。

2.不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月 日

摘要

网络舆情最直观而又最迅速的反映了社会各层面的社情状况与发展形势，其作为社情民意中最广泛和最犀利的一部分对社会产生的影响越来越广泛。但是，互联网的快速发展也愈体现出了从人控制一切工具，向摆脱人的控制、继而控制人的特殊物质方向发展的态势。它在给人们提供服务和便利同时，负面影响也开始慢慢的凸显。鉴于此，为了正确引导公众的舆论走向，建立一套适用于我国政府的网络舆情监测势在必行。

论文以网络舆情监测为具体研究对象，采用文献综述和理论服务于实践的研究方法，分五大部分展开网络舆情监测系统的设计工作。首先介绍研究状况，并阐述了论文的研究意义；其次主要阐述相关技术的理论综述，为下文的分析奠定理论根基；再次基于当前网络舆情的发展现状基础上，展开总体分析，并提出系统构建的需求分析；基于需求分析结果，采用 B/S 架构对系统展开总体架构设计，并依据设计展开功能性及数据库设计，为系统的实现提高了明确的导向。最后对论文的主要模块给出了实现过程。

本论文的特色在于将 B/S 系统架构与网络信息采集集中进行融会贯通，通过调研需求分析展开系统设计，在一定程度上改变了传统直接搜索引导舆论的运行模式实现了论文的建设目标，为正确引导用户网路舆情提供了便捷。

关键词：网络舆情；垂直元搜索；爬虫技术

Abstract

Internet public opinion the most intuitive and most rapidly reflect social conditions and the development of the situation at all levels of society, the impact of public opinion as the most extensive and most trenchant part of its society more widely. However, the rapid development of the Internet also reflects a more controlled everything from tools to get rid of the control of people, and then a special material to control the direction of human development trend. It is a double-edged sword, in a positive effect, but also brought a lot of negative impact. In view of this, in order to correctly guide public opinion towards the establishment of a suitable network monitoring public opinion, our government is imperative. According to this demand, the dissertation according to the characteristics of the network of public opinion, focusing on aspects related to the refinement of public opinion, the use of effective means to complete the collection and collation of information, so as to realize the concern for social events, to make the correct resolution for the departments and units to provide assistance, showing a good practical guidance and practical significance.

The dissertation describes network monitoring public opinion for a specific study, literature review and theoretical research methods to serve the practice, divided into five parts to start design work network public opinion monitoring system. The first part introduces the research situation and explained the significance of this research; the second part describes the theoretical overview of the related art, laid the theoretical foundation for the following analysis; third development site based in part on the basis of public opinion on the current network, expand the overall analysis and made functional and non-functional requirements analysis system construction. The fourth part based on the result of the need analysis, the use of B / S structure to expand the overall architecture of the system and expand the functionality and design based on database design for the realization of the system to improve the clear guidance. Finally, the paper gives the main module implementation process.

Features of this dissertation lies in the B/S system architecture and network information collected centrally mastery, analysis needs to expand the system design through research, to a certain extent, changed the traditional direct search operation mode to guide public opinion to achieve the goal of building paper, correct public

opinion to guide users to the Internet provides a convenient.

Key words: Internet Public Opinion; Vertical Metasearch; Architecture; Crawler Technology

厦门大学博硕士学位论文摘要库

目录

第一章 绪论	1
1.1 选题依据及意义	1
1.1.1 选题依据.....	1
1.1.2 研究背景.....	2
1.1.3 实施网络舆情监督的意义.....	3
1.2 国内外研究现状	3
1.2.1 国外研究现状.....	3
1.2.2 国内研究现状.....	4
1.2.3 相关技术发展现状.....	4
1.3 主要研究内容	5
1.4 论文组织架构	6
第二章 系统相关技术	7
2.1 网络信息采集技术	7
2.1.1 垂直元搜索技术.....	7
2.1.2 网络爬虫技术.....	8
2.2 MVC 模式	9
2.3 UML	12
2.4 C/S 模式	12
2.5 本章小结	13
第三章 系统需求分析	14
3.1 可行性分析	14
3.2 业务流程分析	15
3.2.1 数据采集分析.....	15
3.2.2 实时监测分析.....	16
3.2.3 信息筛选去重分析.....	17
3.2.4 系统的服务对象.....	17

3.3 功能性需求分析	18
3.3.1 主要功能需求	18
3.3.2 系统角色需求分析	21
3.4 非功能性需求分析	23
3.4.1 环境及协议需求	23
3.4.2 安全性需求	24
3.4.3 性能需求	24
3.5 本章小结	24
第四章 系统设计	25
4.1 系统设计原则	25
4.2 系统架构设计	27
4.2.1 系统整体架构	27
4.2.2 软件架构设计	27
4.3 系统功能设计	28
4.3.1 系统管理模块	28
4.3.2 数据采集模块	30
4.3.3 信息筛选去重模块	30
4.3.4 统计分析模块	31
4.4 数据库设计	32
4.4.1 设计原则	33
4.4.2 概念结构设计	35
4.4.3 物理结构设计	36
4.5 本章小结	38
第五章 系统实现	39
5.1 系统权限分配	39
5.2 系统网页去噪	40
5.3 系统管理模块	42
5.4 数据采集模块	44
5.5 统计分析模块	45

5.6 系统测试	47
5.6.1 系统界面测试.....	47
5.6.2 系统登录测试.....	48
5.6.3 爬虫规则测试.....	48
5.7 本章小结	48
第六章 总结与展望	50
6.1 总结.....	50
6.2 展望.....	50
参考文献	52
致谢	54

Contents

Chapter 1 Introduction	1
1.1 Research Background and Significance	1
1.1.1 The basis for the selection of the topics	1
1.1.2 Research background.....	2
1.1.3 Significance	3
1.2 Research Status (at Home and Abroad)	3
1.2.1 Foreign research status quo	3
1.2.2 Domestic research status quo	4
1.2.3 Status quo of Technological Development.....	4
1.3 Research Contents	5
1.4 Outline of the Dissertation	6
Chapter 2 System Related Technologies.....	7
2.1 Network information gathering technology	7
2.1.1 Vertical meta search technology	7
2.1.2 Web crawler technology	8
2.2 MVC mode	9
2.3 UML.....	12
2.4 C/S mode.....	12
2.5 Summary	13
Chapter 3 System Requirement Analysis	14
3.1 Feasibility Analysis	14
3.2 Business Process Analysis	15
3.2.1 Data acquisition and analysis	15
3.2.2 Real-time monitoring and analysis.....	16
3.2.3 Analysis of information filtering	17
3.2.4 Service object of the system	17
3.3 Function Requirement Analysis	18
3.3.1 Main functional requirements.....	18
3.3.2 System role demand analysis.....	21
3.4 Non-function Requirement Analysis	23

3.4.1 Protocol requirements.....	23
3.4.2 System Security Requirements.....	24
3.4.3 System Performance Requirements.....	24
3.5 Summary	24
Chapter 4 System Design	25
4.1 Principle of System Design	25
4.2 System Framework Design	27
4.2.1 System architecture	27
4.2.2 Software architecture design	27
4.3 System Function Design	28
4.3.1 System management module.....	28
4.3.2 Data acquisition module	30
4.3.3 Information filter to weight module	30
4.3.4 Statistical analysis module	31
4.4 Database Design.....	32
4.4.1 Design principle.....	33
4.4.2 Conceptual Structure Design.....	35
4.4.3 Physical Structure Design.....	36
4.5 Summary	38
Chapter 5 System Implementation	39
5.1 System authority allocation	39
5.2 System webpage denoising.....	40
5.3 System management module	42
5.4 Data acquisition module.....	44
5.5 Statistical analysis module	45
5.6 System Testing.....	47
5.6.1 System interface test.....	47
5.6.2 System login test.....	48
5.6.3 Crawler rule testing	48
5.7 Summary	48
Chapter 6 Conclusions and Future Work	50
6.1 Conclusions	50
6.2 Future Work.....	50

References.....52
Acknowledgements.....54

厦门大学博硕士论文摘要库

第一章 绪论

1.1 选题依据及意义

1.1.1 选题依据

当前，互联网已广泛应用于社会、政治、经济、科技、文化和军事等领域，给人类信息通讯、舆论传播、文化交流和社会活动带来了巨变，可以说互联网已经完全融入了人们的生活、学习、工作。

我国网民总人数和互联网普及率到 2014 年 6 月份分别达到 6.32 亿和 46.9%，位居世界第一。社会信息传播的主体地位已逐步建立起来，成为公共信息的最大集中和疏散地，公民表达舆情、参与政治最广泛的场所。公众通过网络所表达出来的意见、看法、态度、情绪和行为倾向等就形成了网络舆情。网络舆情作为社会舆情在互联网上的直接反映，因其表达快捷、信息多元、方式互动等优势，倍受广大网民青睐，对国家政治生活秩序和社会稳定产生了深远影响，可是一些网络舆情突发事件，对社会稳定造成了严重威胁，也不得不引起相关部门的重视和思考。

网络舆情最直观而又最迅速的反映了社会各层面的社情状况与发展形势，其作为社情民意中最广泛和最犀利的一部分对社会产生的影响越来越广泛。但是，互联网的快速发展也愈体现出了从人控制一切工具，向摆脱人的控制、继而控制人的特殊物质方向发展的态势。它在给人们提供服务 and 便利同时，负面影响也开始慢慢的凸显，而当负面舆情产生后，带来的负面影响是难以估量的，同时，也是最难以让人能够控制的。互联网对信息传播的快速性，尤其对一些重大新闻和突发事件的报道体现出了传统媒体望尘莫及的优点；同时，互联网具有互动性，这是以往的传统媒体无法相媲美的，民众可以实时的发表自己的真实意见和观点，当一个个网民的意见汇聚到一起便形成了公众意见，继而推动事态的发展和演变，可以说在当今社会，每一个网民都是记者、每一个网民都是摄影师、每一个网民都是新闻编辑；另一方面，由于互联网的匿名特点，一些不法分子或者别有用心的人借助网络散布虚假信息 and 谣言，对人们群众、党委政府都产生了负面影响，并且某些敌对势力也企图通过网络发布各种信息对我国进行渗透和影响和，特别是青年人的思想，从而使他们的价值观念和意识形态得到广泛传播。

1.1.2 研究背景

在当今信息化的时代背景下，以数字化、多媒体、互联网为主流的信息社会，各类新兴媒介传播速度愈发加快。在当今的讯息时代，互联网等新媒体已经成为民众表达各种意见、看法、问题、态度、诉求和参与社会公共事务的平台。舆情就是指公民在自己熟悉的环境中，根据社会事件的出现对社会管理人员所持有的态度，也可以说是，多数公民对社会事件所持情绪、意见、态度、信念等的综合。由舆情的内涵，我们可以知道舆情要通过路径和介质，这种路径和介质可以是文字表达、口头阐述，也可以是个体作为。由于互联网不受时空限制，具有“所有人向所有人的传播”的优势，网络可以将信息瞬间传遍全球的每一个角落，加上网民网上活动自由程度高，可以虚拟身份在网上活动，自由发言，不受约束。今天的人们，越来越多地使用互联网表达意见。无论是信息量，还是观点数，无可争议地压倒了传统媒体。

互联网具有敞开性和虚假性，因此作为舆情重要表达方式的网络舆情包含下列几个方面的特征：

1. 突发性。由于网络敞开性的特点，网络空间的虚拟、隐蔽，对于现实生活和网络世界中的问题，网民可以快速、大胆地上传信息，发表自己的意见。这些意见无需审核把关，一旦通过某一事件而引发的网络舆情会迅速形成，一些情绪化的观点甚至会转变为某种舆情的引火线。

2. 传播广泛性。网络舆情一旦引起网民关注，便会迅速在各类网络载体上传播扩散，让人无法控制。

3. 内容随意性。网络空间法律道德约束较弱，网民在网上可以随意发表言论，嘻怒笑骂基本不受较大约束，随意性极强，事实真相难辩，特别是在网络推手、意见领袖的蓄意炒作下，网络舆情更容易成为被人利用的工具。

4. 偏差性。因为网上发布信息的人员文化水平、道德素质等各方面的信息的不平衡性，其言论和意见不容易受到限制，他们对社会事件和情况所持的态度必然存在差异，以至于网络变成某些人发泄情绪的场所。

5. 难以控制性。现在是舆论主导和舆论影响的时代，网络舆情事件处置往往让人力不从心，一旦形成很难控制。有的会因官方正面回应口径不一、言语不当、表情不适再掀波涛。一些事件因相关人员受到责任追究才最终平息。

6. 直接性。网民可以通过新闻网、博客等网络技术及时而又快速的表达自

己的观点和态度，个人发表的各种言论和意见可以直接上达，无需再经过中间一些不必要的程序。

网络舆情因为其快捷和多元等优势，已经受到广大民众的青睐，对社会秩序和稳定也产生了重大的影响。正确引导、认识和掌控网络舆情，应当成为我们当前必须认真对待的课题。

1.1.3 实施网络舆情监督的意义

近些年，我们国家发生的诸多群体性事件都有一个共同点，那就是与网络有关，不明真相的群众在网上表达不满情绪，然后这些情绪互相影响放大，导致不满情绪的聚集，最后演变成网上网下串联，形成现实危害。因此，网络舆情事件导致了现实危害。这些也不得不让我们思考，如果我们能及时监测网上的热点敏感突发事件，及时了解网民的思想、态度、情绪，并且能够及时的公布政府相关部门的态度、措施，并且积极正面的引导舆情，那么就能将网络更好的为民服务，成为党委政府相关部门治理社会，服务人民群众的新方式、新平台。因此，党委政府需要尽快对处于危机之中的网络舆情进行检测和处理，要抓住解决突发事件的良机。可是当下，由于网络信息渠道多、分布广、数量大，让舆情控制工作显得有心无力。如何快速、准确的筛选出人们关注的热点、敏感、突发信息，从而及时介入、正面引导、快速处置，确保社会大局稳定已经成为相关政府部门必须解决的课题。但仅靠人工用很原始的方法进行收集，显然是跟不上发展的。因此，引入科学的技术方式对舆情进行筛选、掌控显得尤为重要。

1.2 国内外研究现状

1.2.1 国外研究现状

相对比国内而言，国外研究比较早，发展速度也比较快。欧盟委员会自 20 世纪 80 年代起就已经着手对其成员国的民意变化进行监管并检测，希望对他们的方案制定对网络舆情的探索策略分析和工作成果测评起到帮助。美国的有关部门也研究设计了对语言文字进行整合、应用，应对冗杂信息问题的一种新技术。很多高等院校和科研单位进行了深入的合作，一方具有实力雄厚的研究队伍，另一方拥有精尖的技术和领先的设备，可以将无形的科研技术快速的变成有形的商品，通过在实际中的运用来检测科研技术的可行性，并对科研技术的发展形成积极的促进作用。接下来对在全球影响最大的美国 TDT 系统进行简要的介绍。

TDT 系统是美国有关部门设计的对语言文字进行整合、运用，应对冗杂信息

问题的一种新技术。之后，相关技术人员对它作了初期的探索，而且决定用误报率、漏报率对它做出性质与效用方面的评判。后来美国又投入资金启动了一连串的测评会议，为进一步的研究发展明确了方向。奠定了网络舆情监测体系的相关标准。

基本意义而言，TDT 系统包括了话题跟踪、连续文本分割、话题发现、相关发现和新事件发现 5 方面的内容。在体系的发展过程中，要对原有的技术体系进行创新和再创造，为体系发展提供充足的技术基础。

1.2.2 国内研究现状

国内的网络舆情发展相对滞后，由于中西方文化差异显著造成多数的西方先进经验成果不能参照，并且研究部门与应用单位不能得到有效的结合，致使我国的网络舆情商品价格高昂、发展缓慢。不过，从 2006 年开始发展至今，国内也已经有多数高等院校和科研机构对舆情监测系统展开了研究，比较知名的有中科院自动化所、北京理工大学和北大方正技术研究院等。

当然，到目前为止，国内也不缺乏发展比较完备的舆情商品，并且已经应用于实践，对舆情监测系统提供了很大帮助。例：中国科学研究院施行的“天网”就对国家及国际安全、社会稳定、个人及集体的隐私及安全提供了帮助。北大方正推出的智思系统，顺利完成了对繁杂的网络舆情的实时监管与评测，对政府机关凭借传统模式监察评测网络舆情的实行难题进行了很好的处理，强化了对网络内容的监督管理，为国家和管理人员面对突发状况，系统了解民众意见发挥了良好的辅助作用。拓尔思集团的舆情监测体系，在信息的整合和剖析两个角度为舆情监测给予了有效的处理办法。此外，由于国内舆情监测需求的增加，舆情产品数量也出现了迅猛增长。

1.2.3 相关技术发展现状

舆情监测系统是在互联网得到广泛发展的基础上得以出现的，并对其实行监测，特征就是 Web 中的各种数据里自主发觉和获得信息，内容包括语言文字整理、信息搜索、数据发掘和爬虫技术等多个方面，是一项系统性课题，在此主要针对网络信息的采集方面，重点在论坛信息的采集部分做出创新。目前常用的网络信息采集方式主要有：

1. 垂直元搜索技术

垂直元搜索并不是直接从网络搜集信息，而是在原有搜索引擎基础上，整合

Degree papers are in the “[Xiamen University Electronic Theses and Dissertations Database](#)”.

Fulltexts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.