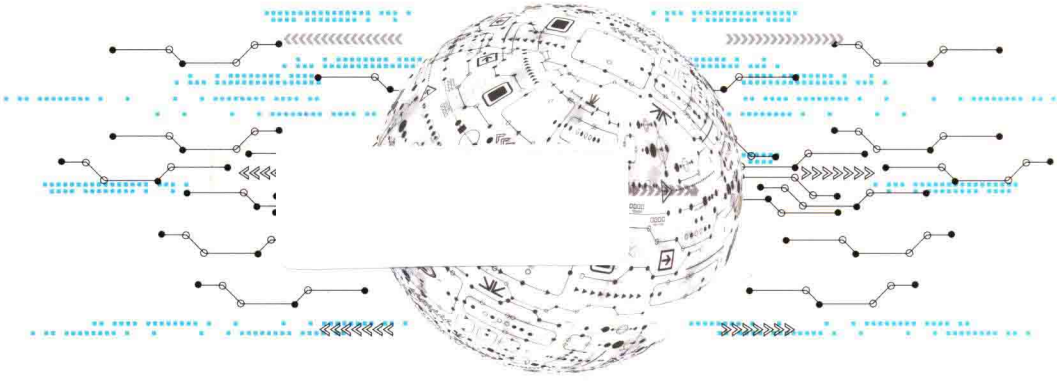
GAOXIAO ZONGHE ZHILI
SHUZIHUA WANGLUO XINXI PINGTAI JIANSHE

高校综合治理
数字化网络信息平台建设

黄斌　郭姣 ▣ 主编



科学出版社

高校综合治理数字化网络信息平台建设

黄 斌 郭 姣 主 编

科学出版社

北 京

内 容 简 介

本书是在广东省财政专项“中医药文化传承创新的大学综治体系及信息技术平台”及“中医药防治重大疾病临床科研信息一体化平台”建设项目的研究成果的基础之上所著。本书主要从理论和技术层面介绍高校综合治理数字化网络信息平台建设。

全书分为六章。第一章综述了高校综合治理网络信息平台建设的意义、发展现状、主要内容及构成。第二、第三章分别是关于校园网络舆情监控技术和校园警情、疫情网络化监控技术的介绍。第四、第五章是社会网络信息分析与决策体系、高校网络信息发布体系的论述。第六章重点介绍防火墙、入侵检测技术、密码技术等校园网络安全技术，本章最后以一个实际案例介绍了高校综合治理数字化网络平台的建设。

本书可作为高等院校综合治理研究人员、信息化建设参与者的参考书，也可为高等院校相关专业的学生和教学人员提供参考。

图书在版编目(CIP)数据

高校综合治理数字化网络信息平台建设 / 黄斌，郭姣主编. —北京：科学出版社，2016

ISBN 978-7-03-048583-0

Ⅰ. ①高… Ⅱ. ①黄… ②郭… Ⅲ. ①计算机网络-应用-高等学校-综合治理-研究 Ⅳ. ①G647-39

中国版本图书馆 CIP 数据核字（2016）第 125243 号

责任编辑：吴卓晶 / 责任校对：陶丽荣

责任印制：吕春珉 / 封面设计：北京睿宸弘文文化传播有限公司

科学出版社 出版

北京东黄城根北街 16 号

邮政编码：100717

http://www.sciencep.com

厚诚则铭 印刷

科学出版社发行 各地新华书店经销

*

2016 年 6 月第 一 版 开本：B5 （720×1000）

2016 年 6 月第一次印刷 印张：9 1/4

字数：185 000

定价：80.00 元

（如有印装质量问题，我社负责调换〈厚诚则铭〉）

销售部电话 010-62136230 编辑部电话 010-62135235

《高校综合治理数字化网络信息平台建设》编写委员会

主　编　黄　斌　郭　姣

副主编　陈沁群　陈　浩　刘来辉

编　委　张洪来　蔡洪民

蔡逸仪　高理文

刘秀峰　骆晓艳

詹秀菊　余萧桓

前　言

构建和谐校园是构建社会主义和谐社会的一个重要组成部分。高等院校（以下简称高校）作为国家知识创新和高层次人才培养的重要基地，其稳定发展具有很强的辐射性、带动性和示范性，对国家安全、社会稳定和经济发展等方面均具有深刻的影响。

随着校园信息化网络建设的不断深入发展，快速高效的校园网在各高校已经初具规模，并逐步朝着教学、科研、管理以及社会服务等功能的网络信息资源共享和整合的一体化方向发展。同时，校园视频监控系统历经三代发展后，基于嵌入式的 Web 网络数字视频监控系统开始逐步成为主流，为构建一个具备面向校园网络安全、环境安全、舆情信息的实时监控能力与互动反馈机制的全方位公共安全网络体系奠定了坚实的物质基础。

目前，在高校维持稳定和治安综合治理工作中，各部门、各系统、各信息平台之间存在各自为政的“信息孤岛”现象，本书从理论和技术层面综合分析现有的机制、系统和信息平台，研究和探索提升高校治安综合治理工作科学化水平的理论和实践。通过数字搜索网络实现对治安视频、网络安全、疫情、舆情等各类信息的统筹监控，对信息进行数据挖掘和统筹分析，及时发现潜在的问题；通过利用多种信息发布渠道，及时开展宣传疏导和指挥调度，最终形成一个信息化和智能化的维持稳定与治安的综合治理快速联动工作机制，以确保高校的和谐稳定发展。

本书由“中医药文化传承创新的大学综治体系及信息技术平台”及“中医药防治重大疾病临床科研信息一体化平台”建设项目的团队成员共同完成。由于编者水平有限，书中难免存在不妥之处，恳请广大读者批评指正。

编　者

2016 年 2 月

目　　录

第一章　概　　述

社会治安综合治理是在党委和政府的统一领导下，在充分发挥政法部门特别是公安机关骨干作用的同时，组织和依靠各部门、各单位和人民群众的力量，综合运用政治、经济、行政、法律、文化、教育等多种手段，通过加强打击、防范、教育、管理、建设、改造等方面的工作，从根本上实现预防和治理违法犯罪，化解不安定因素，维护社会治安持续稳定的一项系统工程。

一、高校综合治理网络信息平台建设的意义

高等院校历来是国家意识形态的前沿领域，国内外各类势力也将高校作为渗透和侵害的重点对象，尤其体现在国家重点、热点领域和敏感时期；同时，高校也是流动人群密集的公开场所，存在着各类传染病疫情突发的公共卫生事件隐患，所以高校的校园稳定、治安安全显得至关重要。高校的安全稳定工作归根到底就是做好校园及周边的社会治安综合治理工作，而社会治安综合治理工作的关键又是以群治群防体系的建设为依托的。

随着信息技术的发展，社会信息化程度的不断提高，治安综合治理工作逐步呈现出其信息化的特征，具体表现在以下两个方面。

（一）信息网络安全对治安综合治理工作的影响比重不断增加

随着高校计算机网络的迅猛发展，网络接入成为当今社会生活中不可缺少的一部分。据统计，中国网民总量占世界第一位，其中31.8%的网民是18～24岁的青年，而在这个群体中，大学生又占据了重要地位。社会焦点问题、偶发事件、关乎学生切身利益的问题，以及具有煽动性的、失实或负面的舆论等极易形成网络舆论，这成为影响校园稳定的重要因素。网络舆情是现实生活中某些热点、焦点问题的反映，其中不乏较强影响力、倾向性的言论和观点。目前，高校网络舆情传播的主要途径有电子公告板（BBS）、博客、微博、电子邮件、即时通信、网络社区等。如果引导不善，负面的网络舆情将对社会公共安全形成较大威胁。从某种程度上讲，高校学生的网络舆情可以说是社会的“晴雨表”和“风向标”。网络舆情监控作为网络管理和网络安全研究中的一个重要部分，越来越引起国家各个信息安全部门的重视，同时在高校维护稳定工作中的地位也显得越发重要。此外，互联网在给人民生活带来便利的同时，网络恶意攻击事件也时有发生，仅在2010年，遭受过病毒或木马攻击的网民比例为45.8%，达到2.09亿人，病毒、木

马、蠕虫、逻辑炸弹等网络攻击事件已经给广大师生带大巨大的危害，网络安全问题也日益成为高校综合治理工作的重要环节。综上所述，信息化的网络平台正逐步成为实现治安综合治理工作跨部门合作、协调、快速响应的有力保障工具。

（二）治安综合治理工作对信息网络技术的依赖不断增加

传统的现场指挥、电话调度等工作模式在应对大规模、分布式群体性公共安全事件时难免出现分身无术、传达不清的尴尬局面；传统的模拟式治安视频监控系统也难以与高度数字化的计算机网络进行有效数据沟通；传统的疫情预报上报制度，如何将海量的各类监控信息组织链接起来，如何对海量的监控信息进行管理、如何在海量的监控数据中查找有用的信息等等，上述种种问题的解决无一不依赖于现代数字化网络技术。因此，大规模的数字化网络监控系统已成为治安综合治理工作顺利开展的前提条件。充分利用即时通信、移动网络、短信、网络视频、博客、微博等各类网络平台对舆情的有效引导和谣言的及时澄清起着重要作用，可以使治安综合治理工作更容易实现对大规模、分布式群体性事件的即时响应、全局把控和精确调度；新一代的基于嵌入式的 Web 网络数字视频监控具有安装部署便捷、节省线缆、轻松实现远程监控访问、可升级性、可扩展性强、可智能处理分析等优势，其正在成为新建、改造或者扩展的安全监控系统项目的首选。此外，电子巡更、智能门禁等数字安防技术在一定程度上也降低了治安综合治理的管理难度。

二、高校综合治理建设的发展现状

（一）各类治安综合治理信息化平台逐步建成，逐步更新

大部分高校已经逐步建成自身的视频监控系统，但很多高校的视频监控系统正处于从模拟向数字过渡的阶段。根据目前趋势，基于嵌入式的 Web 网络数字视频监控系统将开始逐步成为主流。

网络安全系统建设日益完善。从 2008 年开始，国内高校网络安全建设逐步成型，并随着等级保护工作的推进而日趋完善。高校网络安全系统从网络防火墙建设开始，进而逐步建设日志系统、用户接入管理系统、行为审计系统、杀毒软件、入侵检查、漏洞扫描等，到现在通过等级测评及整改建立相对完善的网络安全体系。

舆情监控与引导逐步从现实世界延伸至网络世界。以往通过建立学生信息员和教师信息员队伍，掌握校内舆情动向，并动员辅导员和保卫人员进行疏导；目前，越来越多高校的团委、学生会、保卫处、信息中心等建立了网上交流平台，通过微博、QQ 群等主动与校内师生交流互动，开展网上舆情疏导工作。但是由于经费和技术等原因，大部分高校尚未建设自动化的舆情监控系统，还主要依赖

于公安部门的全网舆情监控系统所提供的通知提示。

电子巡更、职能门禁等信息技术零星建设，尚未普及应用，各高校甚至是同一学校内部的各部门，依照自身的人力和财力资源，开始试探性开展部分信息化管理系统的建设。

（二）各类信息化平台建设相对孤立，缺乏统一的规划与设计，未建立有效的联动机制

治安监控、网络监控和舆情监控相对独立，缺乏有效整合。目前，各高校的治安监控主要由保卫部门负责，网络监控由信息管理部门负责，舆情监控由学生管理部门负责，这三个体系间主要依靠学校治安综合治理办公室协调运作，但其缺乏技术性整合。

各安全系统间相对独立，数据缺乏共享。各类安全系统自身产生大量数据，如视频录像、用户上网日志、防火墙出口日志、病毒报告、舆情信息等，但各类数据缺乏关联，基本处于独立运作或者是事后备查的状态。

各高校尚未建立统一高效的信息化指挥调度中心。目前，各高校应对突发事件的处理方式，主要是依靠现场指挥、会议讨论、电话调度等传统方式进行，这类方式明显无法应对信息化时代信息高速传播模式的需求，信息时代事件从萌芽状态到群体性大规模爆发需要的时间很短，通过短信、即时通信、微博等方式，肇事者可以在极短的时间内大范围地传播非法信息。

（三）校园网络安全检测与舆情控制问题突出

随着计算机网络的普及，网络安全的问题也越来越明显。近年来，频频发生网络攻击事件，这给受害者带来重大损失。此外，由于计算机网络具有交互性特点，人们不再像过去一样仅仅是信息的被动接受者，而是信息的主动传播者，同时以往的传统道德、制度体制、法律行为规范等传统范畴由于受 Internet 的影响也都遭到了一定程度的冲击，Internet 中的负面网络舆情已经成为引发群体性事件的重要因素之一。从云南孟连事件到贵州瓮安事件，再到安徽池州事件、湖北石首事件，大量的群体性事件都与 Internet 有着“不解之缘”，其过程中的诸如谣言、小道消息等负面网络舆情几乎都成了群体性事件爆发的催化剂。随着 Internet 在全球范围内的飞速发展，网络媒体已被公认为是继报纸、广播、电视之后的“第四媒体”。网络信息良莠不齐，其充斥着各种负面的、消极的、虚假的甚至色情的信息，这给大学生的网络环境带来了不良的影响。因此，校园网作为一种典型的局域网，其安全检测与舆情控制问题同样不容忽视。

三、高校综合治理网络信息平台建设的主要内容及构成

（一）高校综合治理网络信息平台建设的主要内容

（1）各类监控系统整合，通过一个统一的监控中心，汇集各类监控系统的信息于一个统一的展示平台，真正达到治安综合治理跨部门、全方位、多渠道的要求。利用设置在校园监控位置的各个信息结点，如视频安防监控点、红外体温非接触测量点，对卫生所、食堂、学生宿舍、图书馆等公共场所的突发事件、盗抢匪情、流感发热等传染性疫情进行收集和定位，记录保存取证。通过网络监控点可以对各种信息交流平台及即时通讯软件（IM）进行监测，从而实现对校园网络舆情有效而全面的掌控。

（2）建立综合的数据分析系统，对各类数据进行关联性的统一分析。各类信息之间都存在一定的关联度，通过建立其联动机制，即可达到牵一发而动全身的效果，真正做到全方位治安综合治理。例如，通过日志信息关联网络结点，通过网络结点关联用户信息和事件地点，通过事件地点关联视频录像，从而形成一体化的信息链。

（3）建立统一高效的信息化综合治理指挥中心。通过整合手机短信、网站信息发布、微博、博客、BBS、校园广播台等通讯渠道，形成一个即时信息发布与交互平台。该平台平时可作为思政宣传、舆情反馈、安全教育、公共卫生信息发布的沟通桥梁，一旦发生公共安全事件，可以作为指挥调度和上传下达的信息平台，也可以积极引导舆论，澄清事实真相，即时回复师生的疑问，迅速疏解矛盾，平息事态蔓延。

（二）高校综合治理网络信息平台的构成

高校综合治理网络信息平台的总体框架如图 1.1 所示，其中，疫情监控系统、治安监控系统、舆情监控系统、网络安全监控系统、日志审计系统用于各类信息的收集取证，综合数据管理平台、数据分析挖掘系统、自动预警系统用于对收集到的信息进行综合、分析和预警；信息综合展示平台、综合信息交互平台则用于信息的不同层面的展示、发布与互动，形成一个校园公共安全信息预警和应急指挥体系。

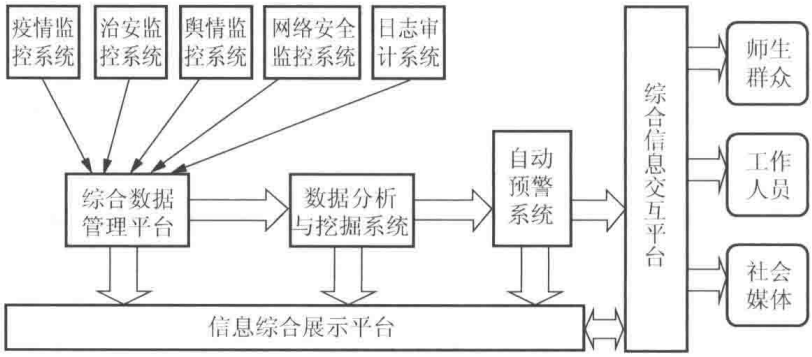


图 1.1　高校综合治理网络信息平台的总体框架

主要参考文献

赵中源，王国栋．2011．高校网络舆情发展趋势及其治理机制[J]．中国高等教育．

中央社会治安综合治理委员会办公室．2009．社会治安综合治理工作读本[M]．北京：中国长安出版社．

中国互联网络信息中心．2011．中国青少年上网行为调查报告[R]．http://www.cnnic.net.cn．

中国互联网络信息中心．2011．第28次中国互联网络发展状况调查统计报告[R]．http://www.cnnic.net.cn．

第二章　校园网络舆情监控技术

所谓舆情，是指在一定的社会空间内，围绕舆情因变事项的发生、发展和变化，作为主体的民众对作为客体的执政者及其所持有的政治取向产生和持有的社会政治态度。

随着互联网使用的日益深入和日常化，近年来网络舆情成为各届人士密切关注的焦点。而莘莘学子聚集的高校校园网，自然是网络舆情较为热力澎湃的地带之一。

一、网络舆情的基本概念

在我国古代，“舆”本意指车箱，代指车。《说文解字·车部》曰：“舆，车舆也。”“舆人”就是造车的工人。《周礼·考工记·舆人》曰：“舆人为车。”至春秋末期，“舆”渐渐演变为轿子的意思，“舆人”也就意指抬轿子的人。随后，其内涵逐渐囊括了差役、小官吏、车夫和随车士卒等所有下层普通大众的意思。至汉代初期，“舆人”与“当芜”和“庶人”一样，成为了普通百姓的代名词。“舆情”一词最早出现于《旧唐书》中的一封诏书，书中称：“朕采于群议，询彼舆情，有冀小康，遂登大用。”舆情在《辞源》（修订本）解释为“民众的意愿”，《现代汉语词典》（第5版）解释为“公众的意见和态度”。可以说，舆情基本上涵盖了民众的情绪、意愿、意见和态度等方面。

舆情的概念，学者们主要从狭义和广义两个方面来研究和表述。例如，王来华认为：舆情狭义上是一种社会政治态度。舆情主体是民众，客体是中介性社会事项，这种政治态度是主体围绕客体的发生、发展和变化对国家管理者而产生的。而程传超则提出了四点阐释：第一，民众和国家管理者之间是利益关系，民众要实现其生存和发展的需要，要依赖于国家管理机构对各种社会利益矛盾进行协调；第二，舆情表面上是民众对中介性社会事项的政治态度，但根本上是对国家管理者的政治态度；第三，中介性社会事项相对“公共事务”而言，都有刺激人们心理的作用，但拓宽了概念的外延；第四，各类有形和无形的因素制约或推着民众社会政治态度在舆情空间中形成和发展。广义上，张克生认为舆情就是社情民意，是民众的外在社会环境和内在主观意愿的总和，主要包括社会客观情况（民情、民力、民智）和主观社会政治态度（民意）。刘毅则认为舆情是情绪、意见和态度的总和。

舆情可以通过民谣、游说和演讲、报纸和杂志、广播和电视以及网络等各种

传播方式来表达。其中，以网络为载体的舆情，则可归为网络舆情的范畴。

当今世界，互联网已渗透到人们生活的方方面面。新闻网站、论坛、微博、微信、QQ、陌陌等，交织成强大的资讯发布、经验分享、情感交流的“天网”。其信息的海量性和流转的迅捷性，超乎想象；对人们的思想意识和价值观念产生了不可估量的影响，特别是对青年大学生的影响。

随着社会的发展和大学生素质的提高，大学生对于政治以及社会事务抱有极大的热情，参与性强，并有着强烈的社会责任感，对于国家事务和社会事务往往有自己独特的看法和主张。网络的飞速发展以及其在高校范围的不断延伸，使大学生的思想交流更加频繁和自由，网络已成为大学生思想交流的最重要的载体。大学生借助网络表达自身对社会事件的看法、态度和意见，在高校网络这一空间内，形成了大学生群体独特的舆情。由此，高校网络舆情就是指在高校网络这一特定空间内，网络舆情主体大学生对自己关心或与自身利益相关的各种社会事务、校园事务所持有的情绪、意愿、态度和意见交错的总和。

鉴于大学生群体的心理特性，高校网络舆情的形成和发展可认为是一种“刺激—反映”的过程。刺激物便是大学生所关注的公共事件或社会突发事件。刺激性信息经过传播为大学生群体所获知，根据这些刺激性信息和大学生自身的知识素养，大学生将自身的态度、意见和建议通过网络这一便捷的渠道表达出来，这类表达的舆情就是大学生对公共事件的反应物。

正确应对校园内的网络舆情，整合现有的 Web2.0 技术和资源来构建网络舆情信息管理模型，对网络舆情进行主动的采集和分析，建立网络舆情预警机制，对网络舆情进行正确的引导和管理，消除网络舆情的不良影响，并建立网络舆情突发事件的快速反应机制，使网络舆情成为推动社会经济政治发展的动力，这在当前具有重要的现实意义。

而从技术层面上，对网络舆情进行监控，主要包括网络舆情信息的收集、网络舆情信息的处理、网络舆情信息智能化分类及预警等几个主要的步骤。下文将一一进行探讨。

二、网络舆情信息的收集

网络舆情信息的收集，主要采用网络爬行器抓取的方式。从程序实质来讲，网络爬行器可以看作一个基于 Web 的多功能高效率的页面扫描程序。网络爬行器可以完成对于 Web 页面完全扫描，在此同时，对 Web 页面进行解析，将其中的超链接加入到等待队列中，作为下一步处理扫描的可能目标。由于超链接在 Web 中拥有广泛应用，因此从理论上分析，网络爬行器在忽略时间的情况下可以对整个 Web 的页面空间进行访问。网络爬行或搜索算法指的是为了保证网络爬行器遍历信息时的深度和广度而制定的一些相关爬行策略。网络爬行器的任务包括：扫

描某个站点获取文件以及文件的层次架构，通过HTML文档访问具体的某个站点，验证超链接的有效性，扫描页面拼写错误等。

（一）网络爬行器的工作原理

网络爬虫工作的流程一般为：爬虫程序从一些起始链接开始，建立 HTTP 通信连接，发送 HTTP 请求，获得响应后保存对应的页面文件，为下一步的工作提供原始的数据。通用网络爬虫在爬行的过程中通常需要维护一个还没有访问链接的集合、一个已经访问过链接的集合和一个不可访问链接的集合。将被网络爬虫处理过的页面对应的链接放到已经访问链接的集合中；如果在爬行的过程中发现无法访问的页面，则将其链接放到不可访问的链接集合中；在获取页面的 HTML 文件以后，通过分析，对其中的链接地址进行提取，过滤掉那些已经访问过或者不能访问或者已经在等待访问的队列中的链接，然后将剩下的链接作为新的等待处理的链接加入到等待访问的链接集合中，爬虫程序不断地重复这个过程直到满足一定的停止条件，结束本次的爬取工作。其工作原理如图 2.1 所示。

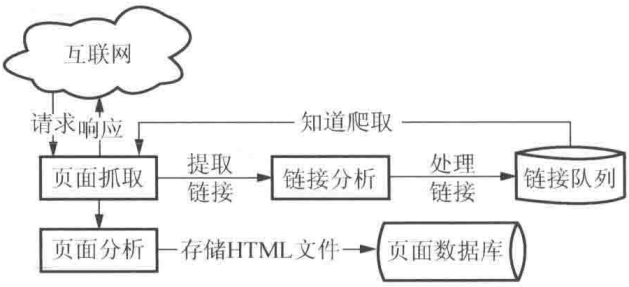


图 2.1　网络爬虫工作原理图

（二）爬行策略分析

广度优先遍历（又称宽度优先遍历）是图论中的经典和常用的遍历算法之一。它是许多重要图论算法的基础，如 Dijkstra 最短路径算法和 Prim 最小生成树算法都使用了这种遍历方法的思想。具体而言，在网页爬行中，广度优先遍历使用了一个先入先出的队列，先对一个页面上所有的链接进行处理，然后才继续处理这个页面所指的下一层页面，直到形成回路或者遍历完成时终止，如图 2.2 所示。这种遍历方式的优点和缺点都比较明显：对于网站的层次比较少的情况，这种遍历方式的收敛速度比较快；但是对于那些规模比较大、网站层次比较多的站点则存在难以深入的问题，对于那些所处位置比较深的页面可能难以进行有效的获取。与广度优先遍历对应的深度优先遍历，则是尽可能深地对图进行遍历。这种方式使用了一个先入后出的队列，深度优先遍历首先从一个页面开始，查看这个页面中是否还有尚未处理的链接，如果有则取出其中一个还没有处理的链接，将其作

为一条爬行的路径，沿着这条路径一直进行下去，直到到达这条路径的终点，即到达一个不含有任何链接的页面或者已经形成回路，然后沿着这条来访的路径回溯，再访问这条路径上其他还没有处理的结点，这个过程一直持续到起始页面上所有的链接都被处理完毕为止。这种遍历方式的优点在于对于那些规模比较大的站点，能够获取到更深的页面，缺点是有可能会造成爬虫的陷入问题，使服务器的压力变大，因此通用网络爬虫一般情况下并不使用这种遍历的方式。

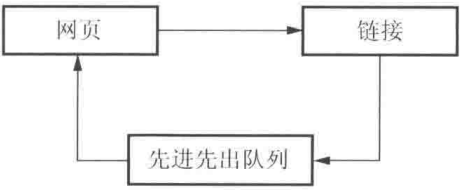


图 2.2　广度优先遍历实现方法示意图

三、网络舆情信息的处理

（一）分词

从网络上抓取而来的原始舆情信息，是杂乱无章的，需要进行一系列的处理。特别地，对于中文舆情原始信息，分词断句，十分重要。

鉴于中文在语言方面的独特性，中文分词与国外应用的分词算法和技术相比，有着较大差别。例如：英语的句子都是由相互间隔的单词组成的，而中文的句子是由没有间隔的单字组成，在对语句进行切分处理时，中文分词明显要难于英文分词。目前，比较流行的分词方法有基于统计的分词法、基于字符串匹配的分词法和基于理解的分词法。

1. 基于统计的分词法

基于统计的分词法是根据相邻两个或多个汉字在中文语句中出现的频率来决定是否划分为一个词的，在实际操作中主要统计其在中文语料库中出现的频率。所以相邻的两个或者多个词在语料库中出现的频率越高，其是一个分词的结果可能性就越大。同时在分词过程中不需要事先准备的词典来指导切词。利用最简单的统计法对“我们都是学生”进行分词。统计法会对“我们/们都/都是/是学/学生”中的分词在语料库中进行统计，找出概率最大的进行最终结果的划分，得到的结果为“我们/都是/学生”。目前，基于统计的分词法有：基于互信息的、基于 N 元文法模型、基于神经网络的和基于隐马尔科夫模型的方法。这是由于统计相关词的计算标准不同造成的，但是不论采用哪种方法，基于统计的分词法有其共同的优点：事先不需要构建庞大的词典，并且具有坚实数学理论支持，分词准确率高。但该方法也有其局限性：由于该方法统计的基础是语料库，故语料库的全面与否

就决定了统计准确性的高低。

2. 基于字符串匹配的分词法

与基于统计的分词法不同，基于字符串匹配的分词法需要一个全面的词典。在分词时，按照某种方式切词，将所切词在字典中匹配，如果有合适的就将该词分出，因此该方法又称为机械分词法。在分词时，可以按照顺序或者逆序切词，也可以按照长度不同，分最大匹配和最小匹配。常用的匹配方式有：正向最大匹配分词、逆向最大匹配分词、最少切分词和双向匹配法。

基于字符串匹配分词法的优点是简单和速度快；但也存在着缺点，即字典词长，严重制约着匹配的速度和准确度。词长时，匹配速度慢；词短时，匹配准确度就会减少。

3. 基于理解的分词法

基于理解的分词法主要以句法分析、语法分析为依据，结合语义分析进行分词。即在分词的同时，要从语料库中的句法、语法结合语义进行分析，模拟人的思维对句子进行理解，在理解的基础上进行分词。该方法处理流程比较复杂，在实际应用中工作量巨大，目前还不能成熟地进行推广。

在完成中文分词的基础上，可以进行语法、句法和语义的分析与处理。有关技术广泛应用于文本的数据挖掘、自动摘要、自动翻译和信息搜索引擎等方面。

（二）网络舆情信息的量化描述

与其他的机器学习过程同理，舆情的量化描述，本质上就是要把文本信息提炼成若干的特征值，从而组成可供模式识别的向量。近年来，众多学者开展了相关的研究。

谢海光等人通过分析某段时间间隔内用户所关注的信息点记录，构建了互联网内容与舆情的热点（热度）、重点（重度）、焦点（焦度）、敏点（敏度）、频点（频度）、拐点（拐度）、难点（难度）、疑点（疑度）、黏点（黏度）和散点（散度）等十个分析模式和判据。

高嘉鑫应用统计原理归纳出 5 个将热门讨论确定为异常事件的相关规则和阈值，并将规则应用到 BBS 进行验证，得出异常事件监测成功率为 100%，准确率为 77%，60%异常事件在 12 小时内即发出通报，最快通报时间为 1 小时内。

郑凌在其硕士论文中利用这个原则进行相关阈值设定和扩展，证明该方法在中大逸仙时空论坛的异常事件监测平均准确率达 75%。研究表明：基于模式识别的舆情监测具有一定的有效性，但由于不同的信息源信息产生规律有较大的差异，该方法具有较大的局限性，只能进行小规模的定点监测。曾润喜等人在发放调查