

网络传播丛书

谢新洲 周锡生 主编

网络数据分析



邱均平 黄晓斌 段宇锋 陈敬全 著

北京大学出版社
<http://cbs.pku.edu.cn>

网络传播丛书

国家自然科学基金资助项目 (70273032)

网 络 数 据 分 析

邱均平 黄晓斌 著
段宇锋 陈敬全

北 京 大 学 出 版 社

• 北 京 •

内 容 提 要

网络数据分析是网络信息计量学 (Webmetrics) 的重要组成部分,也是当前网络界、新闻传播界、信息管理界都十分关注的热点研究领域之一。本书从理论、方法、应用三个角度,全面、系统地论述了网络数据分析的基本原理、方法和工具,详细探讨了它在网络传播、网络信息计量学、网络信息资源管理、电子商务与网络营销、企业管理与市场经营、科学评价与网络管理等许多领域的应用。全书共 9 章,包括网络数据概论、网络数据仓库、网络数据流量分析、网络数据的定性分析、网络数据的多维分析、网络数据的挖掘分析、网络复杂数据的挖掘分析、网络数据分析系统的开发,以及网络数据分析的应用与实例等。

本书既适合于高等院校的新闻与传播、信息管理与科学评价、信息计量学与科学计量学、网站设计与管理、管理科学与工程、电子商务、网络营销等专业的师生教学使用,也可供广大信息工作者、网络工作者、科技工作者和有关管理人员学习参考。

图书在版编目 (CIP) 数据

网络数据分析/邱均平,黄晓斌等著. —北京:北京大学出版社,2004.6
(网络传播丛书)

ISBN 7-301-07371-2

I. 网… II. ①邱…②黄… III. 计算机网络—数据—分析 IV. TP393

中国版本图书馆 CIP 数据核字 (2004) 第 033873 号

书 名: 网络数据分析

著作责任者: 邱均平 黄晓斌 段宇锋 陈敬全 著

责任编辑: 胡伟晔

标准书号: ISBN 7-301-07371-2/TP-0759

出版者: 北京大学出版社

地 址: 北京市海淀区中关村北京大学校内 100871

电 话: 邮购部 62752015 发行部 62750672 编辑部 62765013

电子信箱: xxjs@pup.pku.edu.cn

印刷者: 河北涿县鑫华书刊印刷厂

发 行 者: 北京大学出版社

经 销 者: 新华书店

787 毫米×1092 毫米 16 开本 20.25 印张 414 千字

2004 年 6 月第 1 版 2004 年 6 月第 1 次印刷

定 价: 38.00 元

《网络传播丛书》总序

从1994年到2004年，互联网刚刚在中国走过第一个十年。十年，在人类历史长河中只是短暂的一瞬，但在中国互联网的发展史上，却写下了光辉的篇章！

这十年，是互联网在我国取得突破性发展的十年。据统计，截至2003年11月，我国上网计算机总数已超过3000万台，上网用户人数已达7800万人，并且仍在以每天5万人的速度急剧增长。与此同时，互联网催生的业务像雨后春笋不断出现，互联网不仅为人们拓展出一个获取信息、了解世界、相互沟通的广阔空间，而且正在潜移默化地改变着人们的工作、生活、学习、交往方式，在社会生活的各个方面扮演着越来越重要的角色。

这十年，是互联网作为新兴的第四媒体迅速崛起，在新闻传播领域引发一场广泛而深刻的革命的十年。到目前为止，我国依法取得登载新闻资格的互联网站有150家，全国有1400家新闻媒体创办了网络版。一批由新闻机构创办的综合性新闻网站迅速发展起来，人民网、新华网、中国网、中国日报网、央视国际、国际在线、中青网、中国经济网等重点新闻网站，不仅成为人们获取新闻信息的重要渠道，而且成为其他网站登载新闻的主要来源。重点新闻网站和知名商业网站共吸引了国内95%以上的互联网信息访问量，成为影响我国网上舆论的主要力量。随着互联网的社会影响日益扩大，互联网已经成为我国社会主义新闻宣传体系的重要组成部分。

这十年，是人们对互联网规律的认识不断深化、中国特色互联网新闻宣传体系初步确立的十年。近年来，在党中央的正确领导下，我们顺应时代要求，抓住互联网迅速发展的良机，发挥社会主义集中力量办大事的优势，使得我国互联网新闻宣传事业蓬勃兴起。近年来，我国先后颁布了《全国人大常委会关于维护互联网安全的决定》、《互联网信息服务管理办法》、《互联网站从事登载新闻业务管理暂行规定》等法规和规章，使互联网的发展逐步进入法制轨道。高等院校、学术团体和新闻主管部门的专家学者和新闻工作者撰写了一批传授新知、富有见地的学术著作，在探索建立立足国情、联系实际的中国特色网络传播学方面，进行了不懈努力。可以说，一个以依法管理为基础，中央重点新闻网站为骨干，中央和地方重点新闻宣传网站优势互补，发挥商业网站积极作用的有中国特色社会主义的互联网新闻宣传体系已初步形成。

吹尽黄沙始见金。和所有新生事物一样，中国互联网事业的第一个十年绝非一帆风顺，常常是鲜花伴随着荆棘，掌声夹杂着批评。十年来，互联网经历了投机热潮营造的空中楼阁，也经历了网络泡沫破灭的悲观失望；经历了一些网站片面追求商业利益、刊载虚假错误报道而造成的社会对互联网公信力的质疑，也经历了对互联网媒体属性和社会责任的误

解和分歧。十年过去了，今天我们可以欣慰地说，我国的互联网在经历一个时期的调整之后，开始迎来了它发展的第二个春天。

经过理论的探索和实践的检验，我们对于中国特色互联网事业的性质和发展方向已经有了比较明确的认识：

——互联网站作为社会媒体，在影响社会的同时，也必须承担社会责任，网上传播什么，不传播什么，都有一个主张什么反对什么的问题，都有一个舆论导向的问题；

——互联网站必须成为传播先进文化的重要阵地，要坚持正确的舆论导向，立足于改革开放和现代化建设的实践，着眼于世界文化发展的前沿，发挥网络媒体自身的特点和优势，在内容和形式方面积极创新，大力传播我国优秀传统文化、现代科学技术知识和世界优秀文化成果；

——互联网站必须坚持新闻工作的党性原则，坚持为团结稳定鼓劲和以正面宣传为主的方针，唱响主旋律，打好主动仗；

——互联网站必须纳入法制管理轨道，网上迷信、色情、暴力等有害信息传播，对网民的身心健康和公众利益都造成了很大的危害，对这些问题必须依法进行管理。

中国互联网的第一个十年已经成为历史，第二个十年正在我们脚下展开。积极发展，加强管理，趋利避害，努力使互联网成为思想政治工作的新阵地，对外宣传的新渠道，从而使我国在全球信息网络化的发展中占据主动地位，这是摆在我们面前的一个重要而紧迫的课题。运用好互联网的前提和基础，是探索并掌握互联网传播的特点和规律。摆在我们面前的这套《网络传播丛书》，就是学术界和网络媒体界共同探索的成果。

《网络传播丛书》是北京大学新闻与传播学院联合新华网的几十位作者历时二年多的系统研究成果，是高等院校和新闻宣传一线工作者的共同心血，是集体智慧的结晶。丛书梳理了十年来互联网站，特别是新闻网站的发展历程，深入思考，总结经验，吸取教训，探索规律，对于如何更充分地理解和把握网络媒体的本质和特点，如何对网络传播进行必要的管理和引导，如何改进网络媒体信息的组织与发布方式以提高网络传播的效果，如何为网络传播建立比较科学完整的学科基础和理论体系等问题，进行了有益的探索。研究内容涉及网络媒体的理论、网络媒体的策划与设计、网络媒体的内容分析与利用、网络新闻实务、网络媒体经营与管理、网络广告、网络媒体伦理、网络传播法规与管理等。这套丛书的出版将对我国学术界进一步开展网络传播的研究具有重要的学术价值，也为加强我国网络媒体的应用和普及提供了理论支持。

探索适合中国国情的互联网发展道路，使互联网真正成为先进生产力和先进文化的载体，服务于中国人民的根本利益和中华民族的伟大复兴，需要一代又一代人的不懈努力。互联网是一个新生事物，没有成规可循。让我们共同努力，本着与时俱进的科学精神，在实践与发展中寻求问题的答案。

国务院新闻办副主任 蔡名照

2003年12月

前 言

网络是计算机与通信技术相结合的产物。然而，它的发展和普及更得益于其在社会生活各方面中的不断渗透和日益紧密的结合。在许多场合，被作为网络代名词的国际互联网（Internet），其前身于 1969 年诞生之时，还只是连接了 4 台主机的军用实验网络——ARPANET；然而，到 2003 年 1 月，在网上运行的主机已达 1.7 亿台，网络用户数上升到 6.55 亿户。它已经渗透到社会生活的各个方面，引发了人类社会的巨大变革。网络不仅从根本上改变着现有的生产结构、产业结构、劳动结构，而且也极大地影响着人们的生活方式、交往方式、工作方式、学习方式乃至思维方式，并导致人的价值观和伦理观的深刻变革，一种新型的社会——网络社会应运而生。网络社会、网络经济、虚拟制造、网上教育、电子商务、电子政务、网络医疗、网络伦理、网络文化等与网络相关的一大批新名词、术语的出现，不仅意味着网络对人类社会生活影响的加剧和拓展，同时也表明理论和应用研究领域对网络的关注程度不断增加，这集中反映在与网络相关的研究文献的迅速增长和新的学科分支不断产生等方面。

近几年，在有关网络或文献中出现了两个新的英文术语，即 Webmetrics 和 Cybermetrics，可以意译为“网络信息计量学”或“网上信息计量学”。据目前的文献报道，Webmetrics 是 T.C.阿曼德（T.C.Almind）于 1997 年最早提出来的。1997 年，阿曼德等人在《万维网上的信息计量分析：网络信息计量学的方法探讨》一文中，首次提出了 Webmetrics 这一术语，并认为信息计量学的各种方法完全可以用于万维网上的信息计量分析。关于另一个与之十分相似的术语 Cybermetrics，目前在互联网上已出现了以该词命名的电子期刊或学术论坛，这主要是由西班牙科学信息与文献中心（CINDOC）组织和出版的。

关于“网络信息计量学”的概念，目前国外有几种不同的解释。有的把网络信息计量学定义为是对互联网上的文献进行统计分析的一门学科；有的则把它看成是一门研究互联网上数据之间相互引用的学科；还有的甚至从赛柏空间计算和应用软件的角度，认为网络信息计量学是一门关于计算机软件设计的学科。这些显然都与这门学科的实际情况相差甚远。若从它的研究对象、方法、内容和目标等方面来看，我们认为，网络信息计量学是采用数学、统计学等各种定量方法，对网上信息的组织、存储、分布、传递、相互引证和开发利用等进行定量描述和统计分析，以便揭示其数量特征和内在规律的一门新兴分支学科。它主要是由网络技术、网络管理、信息资源管理与信息计量学等相互结合、交叉渗透而形

成的一门交叉性边缘学科,也是信息计量学的一个新的发展方向和重要的研究领域,具有广阔的应用前景。其根本目的主要是通过网上信息的计量研究,为网上信息的有序化组织和合理分布、为网络信息资源的优化配置和有效利用、为网络管理的规范化和科学化提供必要的定量依据,从而改善网络的组织管理和信息管理,提高其管理水平,促进其经济效益和社会效益的充分发挥。

网络信息计量学是在当前特定的科学背景和技术条件下迅速形成与发展起来的。第一,信息资源电子化、网络化以及网上文献信息数量激增,不仅为网络信息计量学的产生提供了必要的基础和条件,而且还产生了迫切的实际需求,从而推动了这门学科的形成和发展。第二,电子文献信息资料的统计分析及研究成果,为这个学科的形成奠定了基础,积累了经验。早在1997年召开的第63届国际图联(IFLA)大会上,就有3篇论文专门讨论了电子信息资料的统计问题。第三,信息计量学发展的客观需要,随着网上文献信息量的日益增长,信息计量学的研究对象和范围必然要随之扩展到网络领域,这是该学科发展的客观要求和必然趋势。第四,加强和改善网络管理的迫切需要。随着网络化的日益普及,加强网络管理已成为当务之急,而实施定量化管理则是其主要的途径之一。网络信息计量学的研究成果必然会为网络管理的定量化和科学化提供理论指导和定量依据,而网络管理定量化的实践需求又会促进网络信息计量学的全面发展。

我们认为,对网络信息计量学的研究对象应从广义上理解。这里所讲的“网络”,不仅是指互联网,而且也包括局域网等各种类型的网络。网上信息的计量对象主要涉及3个层次或组成部分:

(1) 网上信息本身的直接计量问题,既包括数字信息或文字信息,又涉及集文字、图像和声音为一体的多媒体信息等,如以字节为单位的信息量和流量的计量等。

(2) 网上文献、文献信息及其相关特征信息的计量问题,如网上电子期刊、论文、图书、报告等各种类型的文献,以及文献的分布结构、学科主题、关键词、著者信息、出版信息等的计量,既涉及网上一次文献,又包括二次、三次文献的计量问题。

(3) 网络结构单元的信息计量问题,如网络站点的文献信息量增长、学科分布、信息传递,以及站点之间的相互引证和联系等的计量问题。由此可见,网络信息计量学涉及范围很广,内容非常丰富。与文献计量学和信息计量学相类似,网络信息计量学的内容体系是由它的理论、方法和应用3个部分构成的,而“网络数据分析”则是网络信息计量学的不可缺少的重要组成部分。在这里,对“网络数据”的概念应从广义上理解,不仅是指网上的数字信息,而且还包括网络上的各种符号、数据、信息和知识等。因此,本书涉及的范围非常广泛,内容十分丰富,既研究了网络数据分析的基本原理、主要方法和工具,又探讨了它在网络传播、网络计量、网络评价与管理等许多领域的应用问题。从定性与定量的角度,对海量的网络数据进行全面、系统的分析,无论是对科学技术研究还是经济发展

和社会管理来说都具有重要的理论价值和实际指导意义。

本书是国家自然科学基金资助项目和教育部“十·五”规划项目的重要研究成果之一。我们在长期从事文献计量学和科学计量学的教学与研究工作的基础上,在国内率先开展网络信息计量学及网络数据分析的研究,并在2001年获准主持教育部“十·五”规划项目“网络信息计量学研究(01JA870009)”,2002年又得到国家自然科学基金资助,开展“网络信息计量学的理论、方法与实证研究(70273032)”课题的研究工作,取得了一定进展和较多成果,现借此机会成书出版,既作为《网络传播丛书》中的一种,又是我们近几年来在网络信息计量学,以及网络数据分析方面的项目研究成果的系统总结和升华。在撰写此书的过程中,我们试图从理论、方法、应用3个角度全方位地论述网络数据分析的问题,注重理论与实践相结合,力求使全书的思路清晰、结构合理、内容丰富、观点新颖、资料翔实,既反映和吸收国内外的最新进展,又融入我们自己的系统研究成果,使之具有较强的创新性、科学性、系统性和实用性,既适合于高等院校的新闻与传播、信息管理与科学评价、信息计量学与科学计量学、网络设计与管理、管理科学与工程、电子商务、网络营销等专业的师生教学使用,也可供广大信息工作者、网络工作者、科技工作者和有关管理人员学习参考。

本书由邱均平、黄晓斌主持,首先提出了详细的撰著大纲;然后由黄晓斌、段宇锋、陈敬全、邱均平等分头撰写初稿;最后,邱均平、段宇锋通读全书,做了部分修改和补充,并完成了统稿工作。本书是在作者们近几年来开展网络信息计量学及网络数据分析方面的课题研究的基础上,切磋探讨,互相交流,共同努力完成的,实际上是各位著者的一项集体科研成果。本书出版得到了北京大学新闻与传播学院谢新洲教授和北京大学出版社有关领导的支持和帮助,以及责任编辑胡伟晔同志的辛勤劳动,在此我们表示最诚挚的谢意!

由于我们水平有限,又是多人分头执笔;特别是网络数据分析是一个崭新的专业领域,其研究难度很大;加之时间紧迫,因而书中不妥之处乃至错误在所难免,恳请读者批评、指正。

邱均平于武汉大学

2004年2月

《网络传播丛书》编委会

编委会主任：蔡名照（国务院新闻办公室副主任）

编委会副主任：龚文库（北京大学新闻与传播学院常务副院长、教授）

周锡生（新华网总裁）

谢新洲（北京大学新闻与传播学院副院长、教授）

主 编：谢新洲

周锡生

编委会（以姓氏笔划为序）：

毛 伟（中国互联网信息中心主任）

刘正荣（国务院新闻办公室网络局副局长）

任伟泉（北京北大方正电子有限公司总裁）

肖东发（北京大学现代出版研究所所长、教授）

李伍峰（国务院新闻办公室网络局局长）

罗天文（中国电子科技集团第30研究所所长）

罗以澄（武汉大学新闻与传播学院院长、教授）

陈 刚（北京大学新闻与传播学院广告学系主任、教授）

邱均平（武汉大学中国科学评价研究中心主任、教授）

祝建华（香港城市大学英文及大众传播系教授）

郭庆光（中国人民大学新闻学院院长、教授）

寇晓伟（新闻出版总署音像电子和网络出版管理司副司长）

雷吉成（成都三零信息系统有限公司总经理）

雷跃捷（北京广播学院新闻传播学院副院长、教授）

熊澄宇（清华大学新闻与传播学副院长、教授）

目 录

第1章 网络数据概论	1
1.1 网络与数字信息	1
1.1.1 网络与网络社会	1
1.1.2 网络环境下的数字信息表达	12
1.2 网络数据的类型	20
1.2.1 元数据	21
1.2.2 结构化与半结构化数据	27
1.2.3 非结构化数据	33
1.2.4 流媒体数据	35
1.3 网络数据的特点	38
1.4 网络数据的收集与处理	40
1.4.1 网络数据的来源与收集	40
1.4.2 网络数据的筛选与转换	40
1.4.3 网络数据的组织与集成	41
1.4.4 网络数据的检索与利用	48
1.5 网络数据分析与网络信息计量学的关系	49
1.5.1 网络信息计量学概述	49
1.5.2 网络数据分析与网络信息计量学的关系	50
第2章 网络数据仓库	52
2.1 数据仓库概述	52
2.1.1 数据仓库含义	52
2.1.2 数据仓库的特征	52
2.1.3 数据仓库的构建过程	53
2.1.4 数据仓库的发展	53
2.1.5 数据仓库类型	55
2.1.6 数据仓库的建立	55

2.1.7 数据仓库结构.....	56
2.1.8 数据仓库的开发步骤与实施	57
2.1.9 数据仓库的意义.....	59
2.2 基于网络的数据仓库.....	60
2.2.1 当前数据仓库的局限性	60
2.2.2 基于网络的数据仓库系统的优点	61
2.2.3 基于网络的数据仓库的组建	63
2.2.4 网络数据仓库相关的实现技术	65
2.2.5 基于网络的数据仓库体系结构	67
2.2.6 基于网络的数据仓库面临的问题	70
2.3 数据仓库的典型产品	71
2.3.1 CA的Decision Base.....	71
2.3.2 IBM的DB2 UDB	72
2.3.3 Oracle的数据仓库.....	73
2.3.4 Sybase的Warehouse Studio.....	74
2.3.5 Informix	75
2.3.6 微软SQL Server	76
2.3.7 NCR	77
2.4 数据仓库系统的评价与选择.....	78
第3章 网络数据流量分析.....	80
3.1 概述.....	80
3.1.1 网络基础知识.....	80
3.1.2 网络数据流量分析	82
3.2 网络流量分析的数据来源、指标和技术问题.....	83
3.2.1 网络流量分析的数据来源	83
3.2.2 网络数据流量分析的主要指标	85
3.2.3 网络数据流量分析需要解决的几个问题	88
3.3 网络流量分析的流程和方法.....	90
3.3.1 网络数据流量分析的主要流程	90
3.3.2 网络数据流量分析的主要方法	92
3.4 网络数据流量分析的实施与应用	95
3.4.1 网络数据流量分析的实施	95

3.4.2 网络数据流量分析的应用	99
第4章 网络数据的定性分析	104
4.1 网络数据内容分析	104
4.1.1 内容分析法的产生	105
4.1.2 概念和特点	106
4.1.3 方法和步骤	107
4.1.4 内容分析法的主要类型	110
4.1.5 内容分析在网络数据分析中的应用	111
4.1.6 内容分析的评价	112
4.1.7 内容分析法在网络数据分析中的局限性	113
4.2 网络数据抽象分析	114
4.2.1 抽象思维	114
4.2.2 科学概念	115
4.2.3 科学判断	118
4.2.4 科学推理	119
4.3 网络数据归纳分析	120
4.3.1 归纳分析的概念	120
4.3.2 归纳法的类型	121
4.3.3 归纳分析法在网络数据分析中的作用	124
4.3.4 归纳分析法的局限性分析	124
4.4 网络数据相关分析	125
4.4.1 相关关系的概念及分类	125
4.4.2 网络数据相关分析的步骤	127
4.4.3 相关分析在网络数据分析中的应用	128
4.5 网络数据对比分析	129
4.5.1 对比分析法的概念和理论基础	130
4.5.2 对比分析法的类型	130
4.5.3 对比分析的原则	131
4.5.4 对比分析在网络数据研究中的作用	132
4.6 网络数据分析方法及应用	133
4.6.1 案例节选	133
4.6.2 案例分析	139

第5章 网络数据的多维分析	141
5.1 多维分析概述	141
5.1.1 多维分析的定义	141
5.1.2 多维分析的发展	141
5.1.3 多维分析的特点	142
5.1.4 OLTP与OLAP之间的比较	143
5.2 OLAP的结构	143
5.2.1 OLAP数据的实现方式	144
5.2.2 ROLAP与MOLAP的比较	147
5.2.3 OLAP数据库设计	148
5.2.4 多维分析的基本操作	148
5.3 OLAP的功能	149
5.3.1 对数据的多维观察	149
5.3.2 复杂的计算能力	150
5.3.3 时间智能	150
5.3.4 管理功能	150
5.4 网络数据多维分析工具	150
5.4.1 Microsoft SQL Server的OLAP Services	151
5.4.2 SAS联机分析处理系统	152
5.4.3 DB2多维服务器	153
5.4.4 Hyperion Essbase OLAP Server	155
5.4.5 Crystal Analysis Professional	156
5.4.6 Oracle	156
5.5 网络数据多维分析方法的应用	156
5.5.1 市场和销售分析	157
5.5.2 用户行为分析	158
5.5.3 数据库交易分析	159
5.5.4 预算分析	159
5.5.5 财务报表与整合	160
5.5.6 管理报告分析	161
5.5.7 利润分析	161
5.5.8 质量分析	162

5.6 联机分析的发展趋势——联机数据挖掘	163
5.6.1 数据挖掘与OLAP的区别	163
5.6.2 联机数据挖掘	164
5.6.3 联机数据挖掘模型	165
5.6.4 实现OLAM的机制	166
第6章 网络数据的挖掘分析	169
6.1 数据挖掘概述	169
6.2 数据挖掘过程	171
6.2.1 数据挖掘过程模型	171
6.2.2 网络数据挖掘与知识发现的实现过程	172
6.3 数据挖掘方法	175
6.3.1 统计分析方法	175
6.3.2 遗传算法	175
6.3.3 粗集方法	176
6.3.4 决策树方法	176
6.3.5 神经网络方法	177
6.3.6 模糊逻辑方法	177
6.3.7 聚类算法	177
6.3.8 可视化技术	178
6.3.9 分类方法	178
6.4 数据挖掘的功能	178
6.5 网络数据挖掘	179
6.5.1 网络数据挖掘的内容	180
6.5.2 网络数据挖掘的步骤	184
6.5.3 网络数据挖掘的新技术	184
6.5.4 网络数据挖掘的应用	185
6.5.5 基于网络挖掘的推荐系统	186
第7章 网络复杂数据的挖掘分析	187
7.1 文本数据挖掘	187
7.1.1 文本挖掘的意义	187
7.1.2 文本挖掘特点	187
7.1.3 文本挖掘内容和方法	188

7.1.4 文本挖掘工具	194
7.2 多媒体数据的挖掘	197
7.2.1 多媒体数据挖掘的特点	197
7.2.2 多媒体数据特征的提取	198
7.2.3 多媒体数据挖掘系统的功能模块	200
7.2.4 多媒体数据知识挖掘过程	200
7.2.5 多媒体数据的挖掘方式	201
7.3 时序数据挖掘	203
7.3.1 时间序列分析的特点	204
7.3.2 时间序列聚类分析	205
7.3.3 时间序列建模基本步骤	206
7.3.4 基于时间序列数据的挖掘	206
7.3.5 时间序列分析的应用	207
7.4 空间数据的挖掘	208
7.4.1 空间数据挖掘的含义	208
7.4.2 空间数据的特点	208
7.4.3 空间数据挖掘方法	209
7.4.4 空间数据挖掘系统的体系结构	211
7.4.5 空间数据挖掘处理过程	212
7.4.6 空间数据挖掘的发展方向	213
7.5 网络链接数据挖掘	214
7.5.1 WWW网页链接的结构和类型	214
7.5.2 WWW网页超文本链接的功能和作用	216
7.5.3 网页链接挖掘分析的意义	218
7.5.4 网页链接数据的分析方法	220
7.5.5 网页链接数据的分析案例——6个档案馆网站被链接情况的调查分析	222
第8章 网络数据分析系统的开发	229
8.1 网络数据分析系统设计的基本目标	229
8.2 网络数据分析系统的设计原则	230
8.3 网络数据分析系统的基本功能	230
8.4 数据挖掘分析语言	231
8.4.1 数据挖掘语言的意义	231

8.4.2 数据挖掘语言的设计原则	231
8.4.3 数据挖掘语言的类型	232
8.4.4 基于网络的挖掘语言	233
8.5 网络数据分析系统的结构	236
8.5.1 数据挖掘系统的一般结构	236
8.5.2 基于网络的数据挖掘分析系统	238
8.6 网络数据分析系统的评价	240
8.7 网络数据分析系统工具实例	241
8.7.1 数据挖掘分析工具的分类	241
8.7.2 数据挖掘分析工具的发展	242
8.7.3 数据挖掘分析系统工具的实例	244
第9章 网络数据分析的应用与实例	250
9.1 在网络信息资源管理中的应用	250
9.1.1 国家网络信息资源的建设、宏观控制和管理	250
9.1.2 数字图书馆、网络内容服务提供商的网络信息资源管理	252
9.1.3 企业信息资源管理	254
9.2 在网络行为学研究中的应用	254
9.3 利用网络数据进行在线股票分析	258
9.3.1 利用Excel进行在线股票分析	258
9.3.2 利用专用软件进行在线股票分析	261
9.4 在企业经营及市场分析与预测中的应用	265
9.4.1 企业在线经营分析	266
9.4.2 市场分析和预测	269
9.5 在电子商务与网络营销中的应用	272
9.5.1 电子商务与网络营销概述	272
9.5.2 网络数据分析在电子商务和网络营销中的应用	273
9.6 在网络传播中的应用	278
9.6.1 网络传播的概念	278
9.6.2 网络传播的特点	279
9.6.3 网络传播给信息传播领域带来的影响	280
9.6.4 网络数据分析在网络传播研究中的应用	281
9.7 在网络信息计量学中的应用	282

9.7.1 网络信息计量学概述	282
9.7.2 网络信息计量学的主要研究方法和研究工具	283
9.7.3 网络数据分析在网络信息计量学中的应用	284
9.8 Web of Knowledge及其在科研评价中的应用	290
9.8.1 Web of Knowledge简介	290
9.8.2 Web of Knowledge的特点	291
9.8.3 Web of Knowledge的检索方法	291
9.8.4 Web of Knowledge在科研评价中的应用	296
参考文献	302