新编高等院校信息管理与信息系统专业核心教材

# 网络信息挖掘

# Network Information Mining

黄晓斌 编著

電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

新编高等院校信息管理与信息系统专业核心教材

# 网络信息挖掘

Network Information Mining

黄晓斌　编著

電子工業出版社
Publishing House of Electronics Industry
北京 · BEIJING

**图书在版编目（CIP）数据**

网络信息挖掘/黄晓斌编著. —北京：电子工业出版社，2005.1
新编高等院校信息管理与信息系统专业核心教材
ISBN 7-121-00520-4

Ⅰ.网… Ⅱ.黄… Ⅲ.计算机网络—情报检索—高等学校—教材 Ⅳ.G354.4

中国版本图书馆 CIP 数据核字（2004）第 123433 号

责任编辑：刘宪兰　何　雄
印　　刷：北京牛山世兴印刷厂
出版发行：电子工业出版社
　　北京市海淀区万寿路 173 信箱　邮编　100036
经　　销：各地新华书店
开　　本：787×980　1/16　印张：16　字数：410 千字
印　　次：2005 年 1 月第 1 次印刷
印　　数：5 000 册　定价：24.00 元

凡购买电子工业出版社的图书，如有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系。联系电话：（010）68279077。质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

新编高等院校
信息管理与信息系统
专业核心教材

新编高等院校信息管理与信息系统
专业核心教材

书　目

信息网络技术原理

计算机系统原理

数据仓库与数据挖掘技术

信息系统安全导论

管理信息系统

信息检索导论

信息系统工程

多媒体技术

信息资源开发与管理

数据库技术与应用

电子政务

电子商务原理

数据结构

网络信息挖掘

# 总序

ZONGXU

20 世纪 70 年代，当强大的信息化巨潮还蕴藏在大洋深处，我们的陆地只有一阵微风吹来之时，有识之士们就开始推动信息化专业人才的培养计划，为迎接即将到来的信息化巨潮扩军备战。他们一方面推动着信息技术的普及；一方面根据不同领域的需求，从不同的角度创办了不同类型的信息化专业，这就是管理信息系统专业、经济信息管理专业、科技信息管理专业、医学信息管理专业、林业信息管理专业、农业信息管理专业……实际上，这些专业培养目标可以概括为：为各行业、各部门培养以 CIO 为目标的信息化专门人才。从这一点上看，这些专业的课程设置应当具有相当大的共同性。1996 年，出于多种考虑，教育部将这些专业合并为一个——信息管理与信息系统专业。

以 CIO 为目标的信息化专门人才是一类管理人才。但是他们所管理的主要对象是信息。这样的知识需求，将信息管理与信息系统专业定位于管理学科，与信息学、经济学、法学等学科交叉。这样的学科特点，给课程建设和教材建设带来不少困难。近 30 年来，尽管我们与许多的同行已经进行了不懈的努力，把信息管理与信息系统专业的课程建设和教材建设向前推进了一大步，但是仍然不尽人意，许多课程和教材还没有体现信息管理专业的特色和需要。在多次有关的研讨会上，大家一致呼吁编写一套真正体现信息管理与信息系统专业特色的教材。

新编和出版一套专业教材是要冒风险的，而编写和出版一套以瞬息万变的信息和信息技术为管理对象的专业教材就要冒更大的风险。国内信息业界著名的出版商——电子工业出版社，以超人的胆略愿意同我们一道承担这一风险，组织编写出版一套新的信息管理与信息系统专业核心教材。这套教材冠以“新编”二字，是试图在其体系上能比已有教材更体现信息专业的特色，同时在内容上要能反映最新信息技术的进步以及最新信息管理思想和方法。

目前，国内开设信息管理与信息系统专业的高等院校已经超过 200 所。这样一个数字一方面表明信息化已经深入人心，信息化队伍的规模正在急速扩大，信息化队伍

的素质正在不断提高；另一方面，也给我们增加了巨大的压力，使我们深感责任重大。好在国内本领域的三位知名学者——黄梯云、陈禹、马费成以及其他一批著名专家和后起之秀愿意与我们共担风险，鼓舞了我们挑起这副重担的勇气。同时，我们也把这套教材的不断精化寄希望于广大的同仁，愿我们把这套教材越改越好，永改永新。

新编高等院校信息管理与信息系统
专业核心教材编委会

# 前言

QIANYAN

互联网信息资源越来越庞大，一方面为信息传播开辟了新的途径，另一方面也给用户的利用带来新的挑战。网络信息挖掘是利用现代信息技术与方法，通过构建数据挖掘和知识发现系统，对网络信息进行深入的分析，形成各种有用的知识产品，提供给用户使用，从而实现信息价值的增值。网络信息挖掘是信息管理研究的一个重要内容。通过对网络信息资源的开发、利用和分析研究，可以发现其形成和存在的规律及用户的行为规律，以便合理配置和优化资源，提高网络信息资源的建设和利用水平，从而把信息资源转化为生产力。

本书根据目前网络信息资源管理的发展趋势和完善信息管理专业学生知识结构的需要，对当前网络信息开发和利用技术进行系统的总结和分析，目的是研究网络信息挖掘与利用技术的理论基础，探索有关技术和方法问题，分析在具体应用中的可行性，探讨网络信息挖掘和开发利用的具体途径。本书有如下几个特点：

- 选题新颖，网络信息挖掘是计算机和信息管理领域的前沿课题，书中许多内容都是目前国内外最新的研究成果；
- 内容充实，本书系统分析网络信息挖掘的有关技术原理和规律，重点对网络信息挖掘的方法和技巧等问题进行了探索性研究，并有一定的案例分析；
- 材料比较丰富，书中具有许多实例，并包括各种有关资料和参考书目；
- 适用范围广，本书主要作为信息管理和计算机类本科生和研究生的专业课程教材，也可供广大的信息管理领域科研工作者及其爱好者阅读或参考。

本书按照原理、技术、方法和应用的体例进行编写，共分 9 章：第 1 章介绍了知识发现的基本定义、相关研究的发展和网络信息挖掘与知识发现的意义；第 2 章分析了基于网络的信息挖掘系统的基本要求和功能，知识发现的过程和方法、数据挖掘语言、数据挖掘系统的结构和评价；第 3 章论述了网络信息集成的特点、作用和方式；第 4 章对网络信息的结构挖掘进行了分析；第 5 章对网络信息的内容挖掘进行分析；第 6 章对网络信息的使用记录的挖掘方式方法进行研究；第 7 章讨论了常见的网络信息挖掘的策略；第 8 章介绍了网络信息的挖掘应用的一些实例；第 9 章着重分析网络信息挖掘存在的一些难点和知识发现的研究方向，并对中文网络信息挖掘和知识发现问题提出了一些看法。

本书在编写过程中参阅了国内外一些资料，特向作者表示感谢。电子工业出版社对

本书的出版给予了大力支持，责任编辑刘宪兰和何雄老师对作者给予了极大的鼓励和帮助，并提出了不少意见和建议，其他有关同志也为本书的出版付出了辛勤的劳动，在此表示衷心的感谢。

由于网络信息挖掘是一个新的课题，有许多问题在不断发展变化之中，加上本人水平有限，本书难免会有错漏之处，敬请广大读者批评指正。

黄晓斌

于中山大学康乐园

# 目 录

M U L U