现代信息资源管理丛书

邱均平　主编

信息检索原理与技术

Theory and Technology of Information Retrieval

夏立新　金　燕　方　志　等　编著

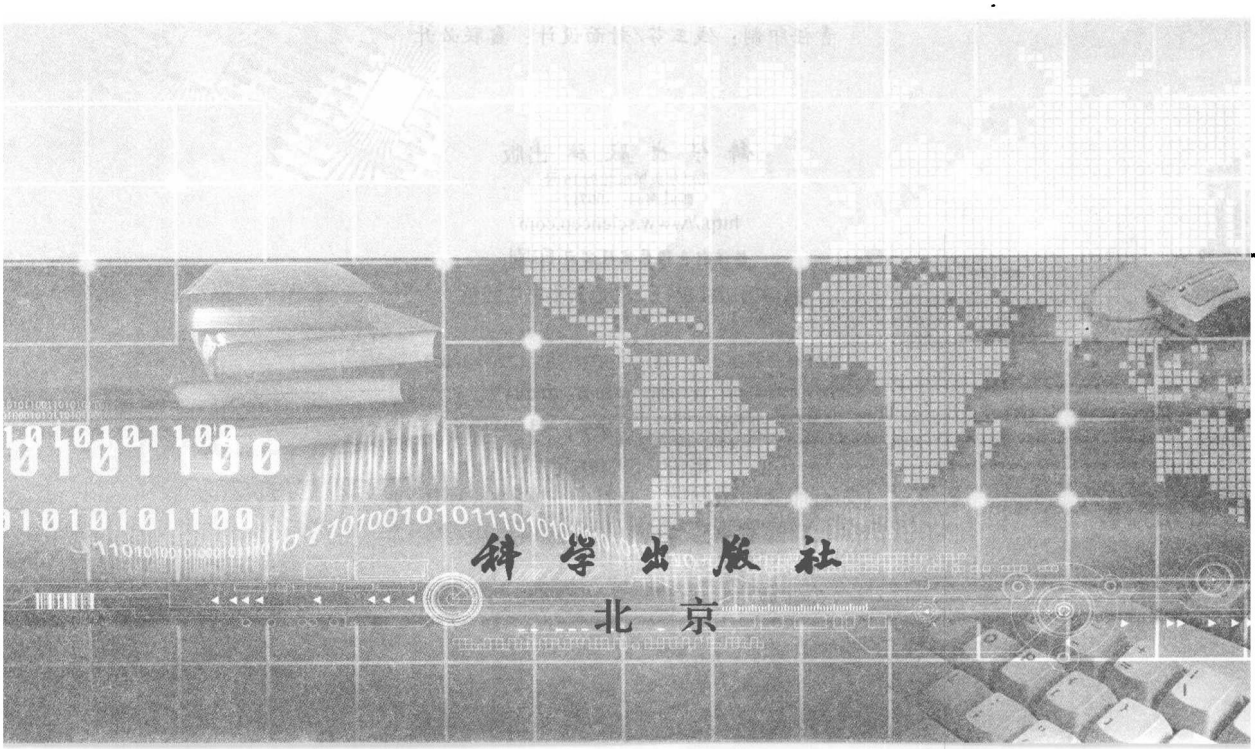
科学出版社
www.sciencep.com

现代信息资源管理丛书
邱均平　主编

Theory and Technology of Information Retrieval

信息检索原理与技术

夏立新　金　燕　方　志　等　编著



科学出版社
北　京

内 容 简 介

本书是《现代信息资源管理丛书》之一。

本书对信息检索领域的相关问题进行了全面系统的研究，既有对其基本理论、方法、技术的论述，也有对其最新发展的系统阐述。具体内容包括：信息检索概论、信息检索模型、自动索引和文档组织、词汇控制、自动文摘技术、用户接口、信息检索系统的评价，以及联机信息检索、因特网信息检索、数字图书馆的信息检索等。

本书可以作为计算机科学与技术、信息管理与信息系统、情报学、图书馆学、档案学等专业的教材或教学参考书，也可供信息中心、情报研究所、图书馆等机构工作人员及广大信息用户学习参考。

图书在版编目(CIP)数据

信息检索原理与技术/夏立新等编著. —北京：科学出版社，2009
（现代信息资源管理丛书/邱均平主编）
ISBN 978-7-03-024870-1

Ⅰ. 信… Ⅱ. 夏… Ⅲ. 情报检索 Ⅳ. G252.7

中国版本图书馆 CIP 数据核字（2009）第 105285 号

责任编辑：李 敏 刘 鹏 / 责任校对：陈玉凤
责任印制：钱玉芬/封面设计：鑫联必升

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

铭浩彩色印装有限公司 印刷

科学出版社发行 各地新华书店经销

*

2009 年 7 月第 一 版 开本：B5（720×1000）
2009 年 7 月第一次印刷 印张：21 1/4
印数：1—3 000 字数：423 000

定价：39.00 元

（如有印装质量问题，我社负责调换〈环伟〉）

《现代信息资源管理丛书》编委会

总　序

信息资源管理（information resource management，IRM）是20世纪70年代末兴起的一个新领域。30多年来，IRM已发展成为影响最广、作用最大的管理领域之一，是一门受到广泛关注的富有生命力的新兴学科。IRM对经济社会可持续发展和提高国家、区域、组织乃至个人的核心竞争力来说，都具有基础性的意义和独特的价值。

在国际范围内，受信息技术进步的推动和经济社会管理需求的牵引，IRM理论研究和职业实践发展迅速，并呈现出一些明显的特征：①广泛融合了信息科学、经济学、管理学、计算机科学、图书情报学等多学科的理论方法，形成以“信息资源”为管理对象的一个新学科，在管理学知识地图中确立了自己的地位。②研究范式的形成和变化。IRM的记录管理学派、信息系统学派、信息管理学派各自发展，以及管理理念、理论和技术方法的交叉融合，形成了IRM的集成管理学派。集成管理学派以信息系统学派的继承和发展为主线，吸收了记录管理学派的内容管理和信息管理学派的社会研究视角，形成了IRM强调“管理”和“技术”，并在国家、组织、个人层面支持决策和各自目标实现的新的研究范式[①]。③研究热点的变化。当前IRM研究在国家、组织、个人层面上表现出新的研究热点，如国家层面的国家信息战略、国家信息主权与信息安全、信息政策与法规、支持危机管理的信息技术等；组织层面的信息系统理论，信息技术（系统）的绩效、价值与应用，IT投资，知识管理，电子商务，电子政务，IT部门与IT员工，虚拟组织，IRM技术等[②]；个人层面的人－机交互、My Li-

① 麦迪·克斯罗蓬．信息资源管理的前沿领域．沙勇忠等译．北京：科学出版社，2005.

② Mehdi Khosrow-Pour. Advanced Topics in Information Resources Management（Volume 1-5）. Hershey：IGI Publishing，2002～2006.

brary、个人信息管理（personal information management，PIM）框架、PIM 工具与方法等[①]。④职业实践的发展。IRM 的基础管理意义和强大的实践渗透力不断催生出新的信息职业、新的信息专业团体和新的信息教育。组织中的 CIO 作为一个面向组织决策的高层管理职位，正经历与 COO、CLO、CKO 等的角色融合与再塑；信息专业团体除信息科学学（协）会、图书馆学（协）会、计算机学（协）会、竞争情报学（协）会、数据处理管理学（协）会、互联网协会等之外，专门的信息资源管理协会也开始成立，如美国信息资源管理协会（Information Resources Management Association，IRMA）；同时，IRM 作为高等教育中的一个专业或课程，广泛渗透于图书情报、计算机、工商管理等学科领域，这种多元并存的教育格局一方面加剧了 IRM 的职业竞争，另一方面也成为推动 IRM 学科发展和保持职业生命力的重要因素。

随着 IRM 在中国的发展，中国的图书情报档案类高等教育与 IRM 的关系日益密切[②]，进入 21 世纪以后，出现了面向 IRM 的整体改革趋势和路径选择。在 2006 年召开的“第二届中美数字时代图书馆学情报学教育国际研讨会”上，与会图书情报（信息管理）学院院长（系主任）签署的《数字时代中国图书情报与档案学教育发展方向及行动纲要》中明确提出：“图书情报档案类高等教育应定位于信息资源管理，定位于管理科学门类”，认为“面向图书馆、情报、档案与出版工作的图书情报学类高等教育是信息资源管理事业健康发展的重要保障”[③]，显示了面向 IRM 已成为中国图书情报档案类高等教育改革的一个集体共识。在这一背景下，图书情报档案类学科如何在 IRM 大的学

① William Jones. Personal Information Management. *See*: Annual Review of Information Science and Technology. Volume 41, 2007.

② 在我国目前的高等教育体系中，图书馆学、信息管理与信息系统、档案学、编辑出版学分别属于教育部高等教育司颁布的《普通高等学校本科专业目录和专业介绍》中的本科专业；图书馆学、情报学、档案学、出版发行学分别属于国务院学位委员会《授予博士硕士学位和培养研究生的学科专业目录》中的二级学科。但它们分别属于不同的学科门类（如本科专业中的管理学类、文学类）和一级学科（如研究生专业中的管理科学与工程，图书馆、情报与档案管理）.

③ 数字时代中国图书情报与档案学教育发展方向及行动纲要. 图书情报知识，2007，(1).

科框架下发展，以信息资源作为对象和逻辑起点进行知识更新与范畴重建，并突出“管理”和“技术”的特点，已成为我国图书情报档案类学科理论研究和教学改革的新的使命和任务。毫无疑问，这将是中国图书情报档案类学科及其教育在新世纪所面临的一次方向性变革和结构性调整，不仅意味着理论形态及其知识体系的改变，也意味着实践模式的革新。《现代信息资源管理丛书》的出版就是出于对这一使命的认识和学术自觉。事实上，我国“图书馆、情报与档案管理”（或称“信息资源管理”）学科领域的教学和研究已经发生了深刻变革，其范围不断扩大，内容更加充实，应用面也在拓展。为了落实“宽口径、厚基础，培养通用型人才”的要求，很多学校的教学工作正在由按二级学科专业过渡到按一级学科来组织，而现已出版的信息管理类丛书仅针对“信息管理与信息系统”专业的需要，适用面较窄，不能满足一级学科的教学、科研和广大读者的迫切需要。因此，根据高等学校 IRM 类学科发展与专业教育改革的需要和图书市场的需求，为了建立结构合理、系统科学的学科体系和专业课程体系，创建符合 IRM 的学科发展和教学改革要求的著作体系，进一步推动本学科领域的教学和科研工作的全面、健康和可持续发展，武汉大学、华中师范大学、黑龙江大学、兰州大学、南京理工大学、中山大学、吉林大学、华东师范大学、湘潭大学、郑州大学、西安电子科技大学和郑州航空工业管理学院等 12 所高校信息管理学院（系、中心）的多名专家、学者共同发起，在广泛协商的基础上决定联合编著一套《现代信息资源管理丛书》（以下简称《丛书》），由科学出版社正式出版。我们希望能集大家之智慧、博采众家之长，写出一套有价值、有特色、高水平的信息资源管理领域的科学著作，既展示本学科领域的最新丰硕成果，推动科学研究的不断深入发展，又能满足教学工作和广大读者的迫切需要。

《丛书》的显著特点主要是：①定位高，创新性强。《丛书》中的每部著作都以著述为主、编写为辅。既融入自己的研究成果，形成明显的个性特色，又构成一个统一体系，能够用于教学；既是反映国内

外学科前沿研究成果的创新性专著，又是适合高校本科生和研究生教学需要的新教材；同时还可以供相关学科领域和行业的广大读者学习参考。②范围广，综合性强。《丛书》涉及“图书馆、情报与档案管理”整个一级学科，包括图书馆学、情报学、档案学、信息管理与信息系统、编辑出版、电子商务以及信息资源管理的其他专业领域，体现出学科综合、方法集成、应用广泛的明显特点。③水平高，学术性强。《丛书》的编著者都具有博士学位或副教授以上职称，都是教学、科研第一线的骨干教师或学术带头人，既具有较高的学术水平和雄厚的科研基础，又有撰写著作的经验，从而为打造高水平、高质量的系列著作提供了人才保障；同时，按照理论、方法、应用三结合的思路构建各种著作的内容体系，体现内容上的前瞻性、科学性、系统性和实用性；在信息资源管理理论与信息技术结合的基础上，对信息技术和方法有所侧重；书中还列举了典型的、有代表性的案例，充分体现其实用性和可操作性；注重整套丛书的规范化建设，采用统一版式、统一风格，表现出较高的规范化水平。

《丛书》由武汉大学博士生导师邱均平教授全程策划、组织实施并担任主编，王伟军、马海群、沙勇忠、王学东、毕强、赵捧未、况能富、范并思、王新才、甘利人、刘永、夏立新、唐晓波、张美娟、赵蓉英、文庭孝、张洋、颜端武担任副主编。为了统一认识，落实分工合作任务，在《丛书》主编主持下，先后在武汉大学召开了两次编委会。第一次编委会（2005 年 11 月 27 日）主要讨论了选题计划，确定各分册负责人。之后，分头进行前期研究、撰写大纲，并报给主编进行审订或请有关专家评审，提出修改意见。经过两年多的准备和研究，2007 年 12 月 23 日召开了第二次编委会，进一步审订了各分册的编写大纲、落实作者队伍、确定交稿时间和出版计划等，并商定在 2008～2009 年内将近 20 本分册全部出版发行。在 IRM 大学科体系框架下，我们选择 20 个主题分头进行研究，其研究成果构成本套丛书著作。这些著作反映了 IRM 领域的重要分支或新的专业领域的创新性研究成果，基本上构成了一个较为全面、系统的现代信息资源管理的学

科体系。参与撰著的作者来自30多所高校或科研院所，有着广泛的代表性。其中，已确定的18本分册的名称和负责人分别是：《信息资源管理学》（邱均平，沙勇忠），《数字资源建设与管理》（毕强），《信息获取与用户服务》（颜端武），《信息系统理论与实践》（刘永），《信息分析》（沙勇忠），《信息咨询与决策》（文庭孝），《政府信息资源管理》（王新才），《出版经济学》（张美娟），《电子商务信息管理》（王伟军），《信息资源管理政策与法规》（马海群），《网络计量学》（邱均平），《信息检索原理与技术》（夏立新），《信息资源管理技术》（赵捧未），《信息安全概论》（唐晓波），《数字信息组织与管理》（甘利人），《企业信息战略》（王学东），《竞争情报学》（况能富），《网络信息资源开发与利用》（张洋）。《丛书》各分册的撰写除阐述各自学科领域相对成熟的知识积累和知识体系之外，还力图反映国内外学科的前沿理论和技术方法；既有编著者的独到见解和新的研究成果，又突出面向职业实践的应用。因此，《丛书》的另一个重要特色是兼具专著与教材的双重风格，既可作为高校信息管理与信息系统、工商管理、图书情报档案、电子商务以及经济学和管理学等相关专业的教材或教学参考书，又可供信息管理部门、信息产业部门、信息职业者以及广大师生阅读使用。

《丛书》的出版得到了科学出版社的大力支持，同时还得到了各分册负责人、各位编著者和参编院校的鼎力帮助；在编写过程中，我们还参阅了大量的国内外文献。在此一并表示衷心的感谢！

由于面向IRM的图书情报档案类学科转型是一个艰巨和长期的任务，我们所做的工作只是一次初步的尝试，不足和偏颇之处在所难免，诚望同行专家及读者批评指正。

邱均平

于武汉珞珈山

2008年6月8日

前　言

信息检索是信息资源管理学科发展非常迅速的重要领域之一。随着因特网的发展及其应用的普及，信息检索理论不断拓展、延伸和丰富；检索手段已经全面更新换代，手工检索方式基本淘汰，取而代之的是基于因特网环境的计算机信息检索；检索技术也得到飞速发展，不再是布尔检索一统天下，而是一个集布尔检索、全文检索、超文本检索、智能检索等多种检索技术综合运用的新时代。在国内外学术界，伴随着信息检索原理与技术研究的不断深入，涌现了不少优秀的学术成果。本书是我们多年来开展信息检索相关问题的研究成果，也是多年来从事信息检索课程教学的经验总结。

本书对信息检索领域的相关问题进行了全面系统的研究，既有对其基本理论、方法、技术的论述，也有对其最新发展的系统阐述，具有重要的学术价值。其具体内容包括：信息检索概论、信息检索模型、自动索引和文档组织、词汇控制、自动文摘技术、用户接口、信息检索系统的评价，以及联机信息检索、因特网信息检索、数字图书馆的信息检索等。

信息检索的研究注重理论与实践的结合。理论方面侧重于论述信息检索的原理、方法及其支撑技术，实践部分侧重于阐述信息资源及其检索工具和方法。现有针对本科层次的信息检索类教材，大多数偏重于信息检索的实务方面，缺乏对信息检索系统的原理、方法和技术的系统全面的论述。与此相比，本书的特点及独到之处如下：

（1）本书系统地介绍了信息检索的原理与技术，具体包括：信息检索概论、信息检索模型、自动索引和文档组织、检索中的词汇控制、自动文摘、检索接口、信息检索系统的评价等基本原理与技术问题，

以及联机检索、因特网信息检索、数字图书馆的信息检索、检索策略等信息检索的方法问题。

（2）本书具有较强的针对性，主要针对信息资源管理类一级学科相关专业的课程教学需要，包括信息管理与信息系统专业、图书馆学专业、档案学专业以及计算机应用等专业相关课程教学的需要。

（3）本书力求突出教学内容的层次性。在内容的编排结构上，借鉴国内外优秀教材的办法，努力做到条理清楚、举例精当、详略得当，对适合于大专院层次、本科生层次及研究生层次的内容进行标注。老师可以根据各个学校的实际情况灵活选用教学内容。

本书的适用对象广泛。从目前国内的专业设置情况看，信息资源管理一级学科涵盖情报学、图书馆学、档案管理三个二级学科，本书适合大专院校信息管理与信息系统专业、图书馆学专业、档案专业等的师生使用。而从信息产业的覆盖范围来看，本书也可以作为信息服务机构的信息服务人员、咨询人员、管理人员的参考书。

本书的顺利出版，首先要感谢本书的合作者——郑州大学信息管理系金燕副教授、陕西理工大学历史文化系方志、武汉理工大学图书馆郭清蓉的辛勤工作。本书的撰写分工如下：夏立新负责写作大纲的设计、全书的统稿，执笔第1、8章，张自然、崔宇奇合作撰写第2章，郭清蓉、张自然合作撰写第3章，王忠义执笔第4章，金燕、代凤明合作撰写第5章，陈晨、夏立新合作撰写第6章，冯小琴执笔第7章，金燕执笔第9章。尤其需要指出的是，中山大学资讯管理系黄晓斌教授对于数字图书馆的信息检索有独到的见解和研究，方志老师根据他提供的全部资料撰写了第10章。韩永青、李小敏两位同学参加了本书第一轮的修改和校对工作，金晶、翟姗姗、李楠、徐晨琛、杜晓曦、李冠楠等同学承担了本书部分章节的文字校对工作。没有我们这个团队成员的通力合作，本书也难以顺利完成。本书还得到了教育部新世纪优秀人才支持计划（项目编号：NCET-08-0788）的资助。

在本书的撰写过程中，我们广泛吸取了国内外有关信息检索原理

与技术的研究成果，参考和引用了大量的相关文献及网上在线资料。特借此书出版之际，我们谨向这些文献作者以及所有关心和支持本书撰写与编辑出版的同志表示由衷的感谢！

信息检索的理论研究在不断深入，信息检索实践应用也如火如荼，本书需要补充和完善的内容还很多，加之作者的水平有限，时间仓促，书中难免出现不妥及疏漏之处，恳请同行专家和读者批评指正。我们期待着能有机会就信息检索原理与技术的有关问题与有志于信息检索领域研究的同仁们进行交流与合作。

夏立新

于武昌桂子山

2009 年 3 月

目　录