

 现代信息资源管理丛书

邱均平 主编

信息资源管理技术

Technology of Information
Resource Management

赵捧未 窦永香 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

Technology of Information Resource Management

信息资源管理技术

赵捧未 窦永香 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。

本书系统简明地论述现代信息资源管理过程中使用的核心技术。全书既注重信息资源管理的技术基础知识的完备性,又强调新方法、新技术的阐述;既突出该领域方法、技术的系统性,又注意提供若干示例的描述。具体内容包括:信息资源采集技术、信息资源组织技术、信息资源压缩与存储技术、信息资源检索技术、信息资源开发与利用技术、信息安全技术等。

本书可作为高等学校信息管理与信息系统、图书情报档案以及相近专业的教材或教学参考书,也可供企业信息管理部门、情报研究所、图书档案馆等专业人士及广大信息用户学习参考。

图书在版编目(CIP)数据

信息资源管理技术 / 赵捧未等著. —北京: 科学出版社, 2010
(现代信息资源管理丛书 / 邱均平主编)

ISBN 978-7-03-029004-5

I. ①信… II. ①赵… III. ①信息管理 IV. ①G203

中国版本图书馆 CIP 数据核字 (2010) 第 182137 号

责任编辑: 李 敏 刘 鹏 赵 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

*

2010 年 9 月第 一 版 开本: B5 (720 × 1000)

2010 年 9 月第一次印刷 印张: 15

印数: 1—3 000

字数: 277 000

定价: 38.00 元

(如有印装质量问题, 我社负责调换)

主编简介



邱均平，武汉大学信息管理学院和教育科学学院教授、博士生导师。华中师范大学特聘教授。我国著名情报学家和评价管理专家、文献计量学的主要奠基人之一。享受国务院政府特殊津贴专家、湖北省有突出贡献的中青年专家。

现任湖北省人文社会科学重点研究基

地——武汉大学中国科学评价研究中心主任、《评价与管理》杂志主编、《图书情报知识》杂志副主编；兼任教育部高等教育教学评估专家，教育部CSSCI指导委员会委员和中国管理科学研究院、南京理工大学等8个单位的研究员、教授或博士生导师，中国索引学会副理事长，中国科学学与科技政策研究会等4个全国性学会的常务理事及《情报学报》、《高教发展与评估》等14种杂志的编委。

一直从事“情报、计量、评价、管理”领域的教学和研究工作，特别在文献计量学、科学计量学与网络计量学、信息管理与知识管理、科学评价与大学评价等方面有精深研究。指导和培养研究生100余名，其学生中不少已成为学术骨干或学科带头人；主持并完成国家和省部级课题28项，获国家社科基金重点项目优秀成果和湖北省社会科学优秀成果省级一等奖（2项）等55项各类学术奖励，特别是近几年来研发的“中国大学及学科专业评价系统”被省级鉴定为“国内领先”成果；出版著作40部，代表作有《文献计量学》、《信息计量学》、《知识管理学》、《大学评价与科研评价》、《中国大学及学科专业评价报告》、《中国学术期刊评价报告》等，其中《文献计量学》首次构建了理论、方法、应用相结合的内容体系，是本学科的奠基之作；《信息计量学》被选为教育部“面向21世纪课程教材”、《知识管理学》被评为“普通高等教育‘十一五’国家级规划教材”；在国内外重要期刊如*Scientometrics*、《情报学报》、《中国图书馆学报》等上发表论文376篇，其中有60余篇获奖或被SCI、SSCI、《新华文摘》、《人大报刊复印资料》全文转载或收录。据权威机构统计和发布，其学术影响力在“图书馆、情报与档案管理”学科领域名列第一，并被收入国际著名的英国剑桥、美国国际《世界名人录》等十多种大型辞书中。

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 etc 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. In: 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,这将是中國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; 然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。会后各分册的撰著工作全面展开, 进展顺利。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》（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

随着国民经济的发展和社会信息化进程的加快,信息资源无论对国家还是对组织机构,都已成为一种战略性资源。信息资源管理也成为一门越来越受到广泛关注和极具生命力的新兴学科。信息资源管理的过程,充分体现出“管理”和“技术”并重的特点,信息技术的不断进步和发展使其研究内容更加充实,应用领域逐渐扩大。因此,系统地思考和剖析并充分利用现代信息资源管理领域中的各种技术,就显得十分迫切和必要。

本书根据《现代信息资源管理丛书》总编提出的基本要求,即建立符合信息资源管理的发展和教学改革要求的学科体系,从学科发展和教学科研的需要出发,系统分析、缜密思考,力求系统而简明地论述现代信息资源管理过程中使用的核心技术。

本书力图从信息资源管理的主要环节出发,用新的视角和新的架构,系统阐述信息资源采集、信息资源组织、信息资源压缩与存储、信息资源检索、信息资源开发与利用和信息安全等主要环节的一些技术。全书既注重信息资源管理的技术基础知识的完备性,又强调新方法、新技术的阐述;不仅突出该领域方法、技术的系统性,还注意提供若干示例的描述,希望使读者达到全面、快速学习的效果。

需要说明的是,本书所涉及的一些内容如信息分析、信息组织、信息检索、信息安全等,属于相对独立的学科领域,但考虑到目前我国信息资源管理领域的研究与教学现状和读者的需要,也选择阐述了这些领域的重要内容。有兴趣的读者可进一步参阅专门的书籍。

本书的结构如下。

第1章介绍了信息资源、信息资源管理的入门知识,并扼要阐述

信息资源管理的信息技术基础,如通信技术、计算机技术、人工智能技术。

第2章介绍了信息资源采集技术,包括文本信息采集、图像信息采集、音频信息采集、视频信息采集等技术,同时还简要介绍了近年来采用的射频识别、无线传感器网络、电子标签等新技术。

第3章讲述了信息资源组织技术,包括标引、聚类与分类、信息摘要技术、信息资源内容描述等信息资源组织体系的基础内容,同时简要论述了体现信息组织综合性和结构性的信息构建技术。

第4章介绍了信息资源压缩与存储技术,主要介绍信息压缩的概念和方法以及文本、图像、音频、视频等信息的主要压缩技术,同时阐述信息存储的基础技术以及存储技术的新发展。

第5章论述了信息检索技术。基于信息资源管理领域对信息检索技术的基本知识需要,主要论述文本信息检索技术、多媒体信息检索技术、并行与分布式检索技术、跨语言检索技术、智能检索技术、自然语言检索技术和搜索引擎技术等内容。

第6章论述了信息资源开发与利用技术。从信息资源的深层次开发利用的需要出发,系统而扼要地阐述了整个信息资源开发与利用过程中涉及的信息分析技术、数据仓库技术、数据挖掘技术和数字图书馆技术。

第7章讲述了信息安全技术。扼要阐述了信息安全的基本技术,包括密码技术、认证技术、网络安全技术以及数据库安全技术。

本书由赵捧未主持编写,负责编写大纲,各章由编写者提出细目并集体研究确定。具体编写任务情况如下:赵捧未编写第1章和第5章,窦永香编写第2章和第3章,刘怀亮、刘成山编写第4章,秦春秀编写第6章,刘成山、陈希编写第7章,最后由赵捧未、窦永香修改统稿。赵飞、淡金华、邓婕、翟小静、马超、江璐、李春燕、史博文、赵丽彬、陈静等多名研究生参与了一些章节的资料准备、编写与修改工作。

本书的编写得到了丛书总编邱均平教授的直接指导和多位编委的大力帮助。

在全书的编写过程中，参阅了大量的国内外有关专家学者的著作、论文和教材。限于丛书的板式设计等原因，我们未能以脚注一一对应标注，而是以参考文献的形式列于书末。这些作者的论著为本书提供了丰富的素材，在此表示衷心的感谢！

信息资源管理是一个综合性的学科领域，所涉及的技术也处在不断发展之中，加之类似的书籍尚不多见，因此本书的编写也是一次尝试，其体系结构、内容选择等方面一定有诸多需要改进和完善之处，恳请读者批评指正。

赵捧未

2010年7月于西安

主要作者简介

赵捧末，博士，教授。现任西安电子科技大学经济管理学院院长，西安电子科技大学“管理科学与工程”学科带头人。兼任教育部管理科学与工程学科专业教学指导委员会委员、中国国防科技信息学会常务理事、国际信息系统协会中国分会常务理事、中国电子学会情报分会副主任委员、陕西省电子商务与电子政务重点实验室学术委员会委员等职务。

主要从事信息资源管理、信息组织与检索、数字图书馆等领域的研究与教学工作。主持国家自然科学基金项目、部委省级研究课题多项，在国内信息管理、情报学图书馆学核心期刊及重要国际学术会议发表论文60多篇。获得陕西省优秀教学成果一等奖1项；主讲的“信息管理学”是陕西省精品课程，负责建设的本科专业“信息管理与信息系统”是陕西省名牌专业和特色专业。

目 录

总序

前言

第1章 概论	1
1.1 信息资源	1
1.1.1 信息资源的概念	1
1.1.2 信息资源的类型	2
1.2 信息资源管理	3
1.2.1 信息资源管理的含义	3
1.2.2 信息资源管理过程	3
1.3 信息技术基础	4
1.3.1 通信技术	4
1.3.2 计算机技术	6
1.3.3 人工智能	9
第2章 信息资源采集技术	12
2.1 文本信息采集技术	12
2.1.1 计算机传统输入技术	12
2.1.2 语音识别技术	15
2.2 图像信息采集技术	17
2.2.1 扫描仪	17
2.2.2 数字照相	19
2.2.3 条形码技术	21
2.3 音频信息采集技术	24
2.3.1 音频采集	24
2.3.2 波形声音的采集、处理和输出	25

2.3.3 语音合成	25
2.4 视频信息采集技术	26
2.4.1 模拟摄像机和视频采集卡	26
2.4.2 数码摄像机	27
2.5 信息采集的新技术	27
2.5.1 射频识别技术	28
2.5.2 无线传感器网络	29
2.5.3 电子标签	32
第3章 信息资源组织技术	33
3.1 标引	33
3.1.1 自动标引	34
3.1.2 基于词汇分布特征的标引方法	37
3.1.3 基于语言规则与内容的标引	38
3.1.4 人工智能标引方法	39
3.2 聚类与分类技术	42
3.2.1 常用聚类方法	42
3.2.2 常用分类方法与技术	44
3.3 信息摘要技术	49
3.3.1 文本信息摘要的生成与实现技术	49
3.3.2 网页信息摘要的生成与实现技术	50
3.3.3 视频信息摘要的生成与实现技术	53
3.4 信息资源内容描述	56
3.4.1 元数据	56
3.4.2 超文本标记语言	66
3.4.3 可扩展标记语言	68
3.4.4 资源描述框架	71
3.5 信息构建	76
3.5.1 信息构建的含义	76
3.5.2 网站 IA	77
3.5.3 信息构建的应用	80

第4章 信息资源压缩与存储技术	82
4.1 信息压缩技术	82
4.1.1 数据压缩方法的分类	82
4.1.2 数据压缩技术的性能指标	83
4.1.3 常用的数据压缩方法	84
4.1.4 文本信息压缩技术	86
4.1.5 图像信息压缩技术	86
4.1.6 音频信息压缩技术	89
4.1.7 视频信息压缩技术	90
4.2 信息存储技术	93
4.2.1 信息的印刷存储技术	93
4.2.2 信息的磁存储技术	94
4.2.3 信息的激光存储技术	96
4.2.4 信息的半导体存储技术	99
4.2.5 信息存储技术的新发展	101
第5章 信息检索技术	103
5.1 信息检索的含义与数学模型	103
5.1.1 信息检索的含义	103
5.1.2 信息检索数学模型	104
5.2 文本信息检索	106
5.2.1 顺排文档检索	106
5.2.2 倒排文档检索	109
5.2.3 加权检索	112
5.2.4 全文检索	114
5.3 多媒体信息检索	116
5.3.1 多媒体技术	116
5.3.2 多媒体信息检索技术	117
5.4 并行与分布式检索	119
5.4.1 并行信息检索	120
5.4.2 分布式信息检索	124
5.5 跨语言检索	128

5.5.1 跨语言检索基本概念	129
5.5.2 跨语言检索相关技术	129
5.5.3 跨语言检索实现策略	131
5.6 智能检索	135
5.6.1 智能信息检索的概念和特点	136
5.6.2 智能信息检索的系统结构	136
5.6.3 智能信息检索原理	137
5.6.4 智能信息检索的核心技术	137
5.6.5 智能信息检索的主要方法	139
5.7 自然语言检索	141
5.7.1 基于语法分析的自然语言检索	141
5.7.2 基于语义分析的自然语言检索	142
5.7.3 基于本体的自然语言检索	143
5.8 搜索引擎技术	144
5.8.1 搜索引擎的基本概念与分类	144
5.8.2 搜索引擎技术原理	146
5.8.3 搜索引擎系统的发展趋势	147
第6章 信息资源开发与利用技术	149
6.1 信息分析	149
6.1.1 信息分析的概念	149
6.1.2 信息分析的特点	150
6.1.3 信息分析的方法	153
6.1.4 信息分析的常用技术	163
6.2 数据仓库	169
6.2.1 数据仓库的含义和特点	169
6.2.2 数据仓库的体系结构	171
6.2.3 数据仓库的关键技术	173
6.2.4 数据仓库的支撑技术	176
6.3 数据挖掘	178
6.3.1 数据挖掘的含义和任务	178
6.3.2 数据挖掘技术	180