

 现代信息资源管理丛书

邱均平 主编

数字信息组织

Organization of Digital Information

甘利人 薛春香 刘 磊 等 编著



科学出版社

www.sciencep.com

现代信息资源管理丛书

- 信息资源管理学
- 信息资源管理技术
- 信息资源管理政策与法规
- 政府信息资源管理
- 网络信息资源开发与利用
- 信息检索原理与技术
- 信息系统理论与实践
- 信息咨询与决策
- 信息获取与用户服务
- 信息安全概论
- 信息分析
- **数字信息组织**
- 数字资源建设与管理
- 企业信息战略
- 电子商务信息管理
- 竞争情报学
- 网络计量学
- 出版经济学

ISBN 978-7-03-026446-6



9 787030 264466 >

销售分类建议：信息管理

定价：38.00元

 现代信息资源管理丛书

邱均平 主编

Organization of Digital Information

数字信息组织

甘利人 薛春香 刘 磊 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。

本书立足于数字信息特点,全面深入地阐述了数字信息组织的理论、方法、技术及应用,内容涉及数字信息及数字信息组织基本内容与方法、分类法与主题法、语义网信息组织与本体、数字信息组织描述规范、数字信息组织技术标准、数字信息自动组织技术、数字图书馆信息组织、网站信息构建、电子商务网站信息组织和电子政务信息组织等。本书侧重于网络环境中数字信息组织内容、方法、技术的系统论述,具有系统性、新颖性、较高的学术性和较广泛的应用价值。

本书可作为高等院校信息管理与信息系统、图书馆学、情报学等专业的教材或教学参考书,也可供各类从事与信息组织相关工作的机构或部门,包括各类型的文献机构、网络公司及企业或政府信息管理部门的工作人员学习参考之用。

图书在版编目(CIP)数据

数字信息组织 / 甘利人等编著. —北京: 科学出版社, 2010

(现代信息资源管理丛书 / 邱均平主编)

ISBN 978-7-03-026446-6

I. 数… II. 甘… III. 计算机应用 - 情报检索 - 高等学校 - 教材
IV. G252.7

中国版本图书馆 CIP 数据核字 (2010) 第 010816 号

责任编辑: 李 敏 刘 鹏 / 责任校对: 朱光光

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>

铭浩彩色印务有限公司印刷

科学出版社发行 各地新华书店经销

*

2010 年 1 月第 一 版 开本: B5 (720 × 1000)

2010 年 1 月第一次印刷 印张: 20

印数: 1—3 000

字数: 400 000

定价: 38.00 元

如有印装质量问题, 我社负责调换

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

《现代信息资源管理丛书》编委会

主 编 邱均平

副主编 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

编 委 (以姓氏汉语拼音为序)

毕 强	常金玲	陈 远	程 妮
邓香莲	窦永香	段宇锋	范并思
付立宏	甘利人	黄晓斌	金 燕
况能富	刘 永	刘焕成	罗 力
罗贤春	吕元智	马海群	马瑞敏
牛培源	邱均平	沙勇忠	苏金燕
索传军	谭必勇	谭春辉	唐晓波
汪传雷	王桂萍	王伟军	王新才
王学东	王应解	王曰芬	文庭孝
夏立新	夏义堃	肖秋惠	肖仙桃
薛春香	颜端武	杨 峰	余以胜
张 蕊	张 洋	张美娟	赵捧未
赵蓉英	朱少强	邹 瑾	

秘 书 余 波

前 言

随着现代科学技术的发展,人类社会从信息社会向知识社会迈进,信息资源的开发日益受到广泛的重视。计算机技术、通信技术和网络技术的飞速发展,使得数字信息呈指数增长,“失去控制和无组织的信息不再是一种资源”。因此,在数字化、网络化这样新的信息环境中,数字信息的组织管理是必需且迫切的。

数字环境中的信息组织问题已提上议事日程,关于这方面的研究也可谓轰轰烈烈,比如知识组织、元数据、信息构建、数字图书馆、电子商务信息组织、电子政务信息组织等都是该领域当前的研究热点。目前,图书情报档案专业的业务平台正逐渐从传统环境转向数字环境。因此,如何将传统的信息组织理论与实践跟当代全新的数字信息组织需求及研究成果结合在一起,形成一个面向数字环境的信息组织知识体系这一问题,无论是对图书情报档案学科在数字时代寻找新的发展契机,还是对信息管理与信息系统、图书馆、情报和档案管理专业学生掌握时代发展所需要的知识技能,都具有重要意义。

本书给出了数字环境中信息组织的概念、内容、方法、技术支持体系以及应用的整体框架,并且在继承传统信息组织最精华内容的同时,充分注重将分散在各个领域中最新的有关信息组织的研究主题融为一个有机的整体,以各类信息服务网站为基本平台,提供上述信息组织理论的研究成果、应用案例,为本科生传授“看得见,摸得着”的实用知识与技能。全书共分10章,分别从理论与方法、技术与规范、应用与实践三个层面来论述,具体内容如下:

第1章为绪论,概述了数字信息的概念、特点和类型,简述了数字信息组织的概念、内容、特点、方法和技术手段,并总结了数字信息组织的发展历史和意义。

第2章概述了数字信息组织传统的语义工具——分类法与主题法,并着重阐述文献分类法和叙词表在数字信息组织中的应用。

第3章结合语义网环境,介绍数字信息组织的新型语义工具——本体以及本体的基本内容、构建方法和构建工具,并结合实例讲解本体的具体构建步骤。

第4章阐述了数字信息组织的描述规范——元数据,重点介绍了元数据描述体系和基于XML/RDF的元数据标记。

第5章着重讨论了数字信息组织的各种技术标准,包括ISO标准、W3C标准以及我国的国家标准。

第6章重点描述了数字信息组织自动化的相关技术,包括自动采集、自动标引、自动分类和聚类以及信息服务的相关技术。

第7章阐述数字图书馆中信息组织和管理的主要问题,侧重论述数字图书馆中分布式资源互操作的实现。

第8章从网站信息构建的角度阐述基于网站的信息组织,包括信息构建的基本要素、基本方法,并阐述如何对网站用户可用性进行测试来评价数字信息组织的效果。

第9、10两章分别阐述了面向不同数字对象的信息组织,包括电子商务网站和电子政务网站的信息组织。电子商务网站信息组织重点论述如何基于信息构建理论来实现网站规划,提高网站的用户可用性;电子政务网站信息组织侧重政务实体信息之间的交互和应用。

本书为集体智慧和劳动的成果,主要由南京理工大学信息管理系和南京农业大学信息管理系从事相关教学的教师共同完成。本书由甘利人、薛春香、刘磊主编,甘利人、薛春香、刘磊、岑咏华、丁晟春、冯英华、胡以涛、吴鹏、颜端武、章成志参加了撰著工作。甘利人负责拟定大纲及定稿,各章分工如下:第1章由刘磊、冯英华撰写,第2章由薛春香撰写,第3章和第9章由丁晟春撰写,第4章和第10章由吴鹏撰写,第5章由刘磊和胡以涛撰写,第6章由章成志撰写,第7章由颜端武撰写,第8章由岑咏华撰写。薛春香做了大量统稿和审校工作,西安电子科技大学刘玮为相关章节编写提供了宝贵资料。

数字信息组织是一个很大很新的研究领域,观念更新和技术发展都比较迅速,既受传统文献组织方法和手段的影响,又有新型技术的突破。作者不揣浅陋,试图将两者结合,并侧重分析新型技术在数字信息组织中的应用。由于多人执笔,且很多理论和方法尚未形成定论,

内容上难免有疏漏、不妥之处，希望各位读者不吝赐教。

本书得以及时完成和顺利出版，首先，要感谢丛书总编邱均平教授的大力指导和鞭策；其次，感谢本书参考和引用的论著作者，是他们的研究成果为本书的撰写提供了宝贵资料和启示。

编著者

2009年8月于南京

目 录

总序

前言

第1章 绪论	1
1.1 数字信息概述	1
1.1.1 数字信息的概念	1
1.1.2 数字信息的特点	2
1.1.3 数字信息的类型	3
1.2 数字信息组织概述	6
1.2.1 数字信息组织的概念	6
1.2.2 数字信息组织的基本内容	6
1.2.3 数字信息组织的特点	7
1.3 数字信息组织的方法体系	8
1.3.1 数字信息组织的基本方法	8
1.3.2 数字信息组织的语义工具	9
1.4 数字信息组织的规范标准、主要技术与管理模式	11
1.4.1 数字信息组织的描述规范与技术标准	11
1.4.2 数字信息组织的主要技术	12
1.4.3 数字信息组织的管理模式	14
1.5 数字信息组织的历史与意义	15
1.5.1 信息组织的历史沿革	15
1.5.2 数字信息组织新问题的提出	21
1.5.3 数字信息组织研究的意义	22
1.6 本书知识框架	23
习题一	25
第2章 分类法与主题法	26
2.1 分类法概述	26

2.1.1	分类法的原理	27
2.1.2	分类法的类型	29
2.1.3	分类法的作用	31
2.2	数字信息的分类组织	31
2.2.1	数字信息分类的意义	31
2.2.2	数字信息的分类标引	32
2.2.3	数字环境中的分类法	33
2.2.4	数字信息分类体系构建	39
2.3	主题法概述	44
2.3.1	主题法的类型	44
2.3.2	主题法的特征和作用	45
2.3.3	主题法的原理	47
2.4	数字信息的主题组织	51
2.4.1	数字信息主题组织的意义	51
2.4.2	数字信息的主题标引	51
2.4.3	数字环境中的主题法	52
2.5	从传统语义工具到知识组织系统	56
2.5.1	知识组织系统的概念及由来	56
2.5.2	知识组织系统的类型	57
2.5.3	知识组织系统的特征和作用	58
习题二		60
第3章 语义网信息组织与本体		61
3.1	语义网	61
3.1.1	语义网的架构	61
3.1.2	语义网环境中的信息组织	65
3.2	本体概述	66
3.2.1	本体的内涵及发展	66
3.2.2	本体的构成	67
3.2.3	本体的类型	68
3.2.4	本体与分类法、叙词表的比较	69
3.2.5	本体的功能	71
3.3	本体描述语言与构建方法	71
3.3.1	本体描述语言	71
3.3.2	本体构建方法	74

3.4 领域本体的设计	78
3.4.1 本体领域和范围的确定	78
3.4.2 核心概念体系的确定	79
3.4.3 类的属性定义	80
3.4.4 概念间关系的确定	81
3.4.5 实例的创建	84
3.5 领域本体的实现	84
3.5.1 本体构建工具 Protégé	84
3.5.2 基于 Protégé 的协同构建模式	85
3.5.3 基于 Protégé 的领域本体实现	86
习题三	90
第4章 数字信息组织的描述规范	91
4.1 元数据概述	91
4.1.1 元数据的概念与功能	92
4.1.2 数字环境下的元数据特点与类型	93
4.1.3 元数据的研究意义	93
4.2 国内外主要元数据应用体系	94
4.2.1 都柏林核心元数据及其应用	94
4.2.2 其他元数据及其应用	95
4.3 基于 XML/RDF 的元数据标记应用	97
4.3.1 SGML/HTML/XML 标记语言	97
4.3.2 资源描述框架 RDF 概述	104
4.3.3 基于 XML/RDF 的元数据标记应用实例	107
习题四	109
第5章 数字信息组织的技术标准	110
5.1 数字信息组织技术标准概述	110
5.1.1 数字信息组织技术标准研究意义	110
5.1.2 数字信息组织技术标准的主要分类及其功能	111
5.2 国际标准化组织及相关标准	113
5.2.1 国际标准化组织简介	113
5.2.2 ISO 相关标准介绍	114
5.3 互联网联盟及其相关标准	121
5.3.1 互联网联盟简介	121
5.3.2 W3C 相关标准介绍	123

5.4 中国标准化组织及其相关标准	127
5.4.1 中国标准化组织及国家标准介绍	127
5.4.2 相关标准介绍	129
习题五	132
第6章 数字信息的自动组织技术	133
6.1 信息自动采集技术	133
6.1.1 信息采集概述	133
6.1.2 Web 信息资源自动采集技术	134
6.1.3 Google 的“网络爬虫”	135
6.1.4 信息采集的研究热点	136
6.2 信息内容的揭示与标引技术	137
6.2.1 信息内容的揭示与标引	137
6.2.2 信息自动标引技术	138
6.2.3 信息自动摘要技术	142
6.2.4 文本自动聚类	143
6.2.5 多媒体信息内容的揭示与标引	148
6.3 信息服务技术	151
6.3.1 信息检索与信息过滤	151
6.3.2 个性化服务与信息可视化	154
6.3.3 数据挖掘和知识发现	157
习题六	158
第7章 数字图书馆信息组织	159
7.1 数字图书馆概述	159
7.1.1 数字图书馆的概念	159
7.1.2 数字图书馆的特征	160
7.1.3 数字图书馆的研究热点	161
7.2 数字图书馆信息组织的若干基本问题	163
7.2.1 数字图书馆信息组织的原则与方式	163
7.2.2 数字图书馆的体系结构	166
7.2.3 数字图书馆中的信息互操作	171
7.2.4 数字资源的长期保存	177
7.3 国内外典型数字图书馆项目介绍	180
7.3.1 美国数字图书馆项目	180
7.3.2 其他国家数字图书馆项目	182