

21 世纪信息管理丛书

信息组织概论

冷伏海 主 编

徐跃权 史继红 裴成发 顾潇华 副主编



科学出版社

www.sciencep.com

21 世纪信息管理丛书

信息组织概论

冷伏海 主 编

徐跃权

史继红 副主编

裴成发

顾潇华

科学出版社

北 京

内 容 简 介

本书从理论上阐明了信息组织的发展轨迹、基本原则、基本方法和类型,展示了数字信息环境下信息组织的变化、发展趋势和新的信息资源组织方式;探讨了新的信息组织环境下的技术标准、信息组织的分类法、主题法和元数据法的应用和发展;从搜索引擎的信息组织、多媒体的信息组织、EIS中的信息组织和数字图书馆的信息组织的角度概括了网络环境下信息组织的系统建设问题。

本书可作为大专院校信息管理与信息系统、图书、情报、档案等专业教材或教学参考书,也可供信息工作者、图书情报工作者、企事业管理干部学习参考。

图书在版编目(CIP)数据

信息组织概论/冷伏海主编. —北京:科学出版社,2003.7

(21世纪信息管理丛书)

ISBN 7-03-011524-4

I. 信… II. 冷… III. 信息管理-概论 IV. G203

中国版本图书馆 CIP 数据核字(2003)第 043280 号

责任编辑:姚平录 李 敏 / 责任校对:柏连海

责任印制:赵德静 / 封面设计:张 放

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>

源海印刷有限责任公司 印刷

科学出版社发行 各地新华书店经销

*

2003年7月第 一 版 开本:B5(720×1000)

2003年7月第一次印刷 印张:22

印数:1—5 000

字数:430 000

定价:31.00 元

(如有印装质量问题,我社负责调换〈北燕〉)

主编简介

冷伏海，男，1963年8月6日出生于黑龙江省哈尔滨市。1985年7月毕业于北京大学信息管理系，获学士学位，2002年7月毕业于中国科学院文献情报中心，获博士学位。曾任黑龙江大学信息管理学院院长、教授、硕士生导师。现在中国科学院文献情报中心工作。曾主持完成国家、省部级和其他项目15项，发表论文60余篇，出版专著教材10部。2000年获教育部霍英东基金优秀青年教师三等奖。是黑龙江省省级重点学科情报学学科带头人。对商务情报的理论、方法和技术领域有较深的研究。现是国际信息与文献联合会个人会员，国家教育部图书馆学科教学指导委员会委员，中国图书馆学会编辑出版委员会委员和学术委员会教育与培训委员会委员。

《21 世纪信息管理丛书》编委会

主 任 靖继鹏

副主任 柯 平 冷伏海 卢小宾

委 员 (按汉语拼音字母顺序排序)

毕 强 陈忠海 窦平安 李福坤

李景峰 刘 永 马海群 杨文祥

序

21 世纪,是人类社会完成由工业文明向信息文明整体飞跃的伟大时代。在这一历史条件下,社会信息资源的科学组织、建设、管理及有效开发和利用,不仅是人类社会的共同任务,而且是每个社会成员都无法回避的现实问题,上自国家、民族、不同层次的社会组织,下到每一位社会成员,其社会生存和社会发展实力完全系乎于此。因此,信息管理科学的形成与发展不仅具有高度的科学价值,而且具有不可估量的社会实践价值和现实意义。

20 世纪六七十年代,信息管理思想在西方发达国家出现,不久便形成了关于信息和信息资源管理理论的研究热潮。我国在稍后不久也分别从经济信息管理、文献信息管理和信息管理系统等不同的角度开展了这一领域的研究。1998 年 10 月教育部新颁布的《高等院校本科专业目录》在管理科学门类的管理科学与工程学科下设置了信息管理与信息系统专业,从此,信息管理科学的学科建设和专业教育正式进入了我国高等教育领域。这说明,信息管理科学正在从形成阶段走向发展和成熟阶段。可以预期,我国的信息管理科学学科建设和专业教育将在为我国信息化建设提供理论支持和培养高素质信息资源管理专业人才这两方面发挥极为重要的历史作用。因此,我们把信息管理专业教材建设作为一项重要的任务来抓,经过一年多的精心准备,集大家之智慧编写了这样一套有价值、有特色、有创新的适合于信息管理类专业教学的丛书(教材)。

《21 世纪信息管理丛书》由吉林大学工商管理学院院长、博士生导师靖继鹏教授主持,由吉林大学、东北师范大学、黑龙江大学、南开大学、郑州航空工业管理学院、辽宁师范大学、河北大学、郑州大学和山西大学等多所高等院校 40 多位专家、学者合作编写而成。其中,《应用信息经济学》由靖继鹏教授主编;《信息检索》由东北师范大学信息传播与管理学院院长卢小宾教授和山西大学经济与管理学院副院长李景峰副教授主编;《信息系统分析与设计》由郑州航空工业管理学院信息科学系主任

刘永副教授主编;《信息组织》由黑龙江大学信息管理学院院长冷伏海教授主编;《网络信息资源开发与利用》由吉林大学管理学院博士生导师毕强教授和河北大学管理学院杨文祥教授主编;《信息管理概论》由南开大学国际商学院柯平教授和黑龙江大学信息管理学院高洁教授主编;《信息法学》由黑龙江大学信息管理学院副院长马海群教授主编;《电子商务》由吉林大学管理学院信息管理系主任窦平安副教授和辽宁师范大学信息管理系副主任李福坤副教授主编。上述多所院校的相关骨干教师在繁忙的教学中不辞辛劳,为丛书各分册的编写、部分书稿的整理和校对做了大量工作,在此谨致衷心的感谢。

本丛书在满足专业教学之急需的同时,力求全面反映各相关领域国内外最新的研究成果,特别是各院校教学和科研实践中所取得的最新成果。

参与本丛书编写院校的专家、学者自始至终本着“信任、合作、创新、发展”的原则,视推动我国信息管理专业教育发展为己任,充分发扬了良好的团队精神。然而,由于时间仓促,丛书的斧凿之痕可能历历在目。但由于时代的剧变,学术思想和信息技术以令人目眩的速度变化和发展,以及本学科领域教学和科研的急需,不允许我们十年磨一剑,只好将这一套不甚成熟的著作奉献于读者和专家面前。为此,敬望读者与专家不吝赐教,一切不能令人满意之处,俟再版时一并修订。

《21 世纪信息管理丛书》编委会

2002 年 8 月

前 言

信息的组织问题始终是信息管理的核心问题,其理论、方法和技术也始终是信息机构的核心竞争力。无论是在古代的目录学时代,还是近代的图书馆目录时代,以至于现代的图书馆自动化和因特网时代,信息组织一直是在发展着,以应对不同时代对图书、文献、情报和网络化数字化信息的处理。

为了适应网络化、数字化对信息组织的理论与实践的挑战和促进,我们组织一些中青年专业人员撰写了本书。在撰写过程中,我们参考了一些先进国家的该类图书的内容体系和大学课程的内容安排,以期让读者能够了解一些先进国家在信息组织的理论、方法和技术方面的一些进展。

在撰写过程中,我们采取了分工协作的方式。虽然作者分布在国内许多院校,但还是尽量交换资料和信息,以期共享思路和经验,把书写得更耐看。具体的分工是:第1章、第2章、第4章的第1节、第5章的第6节和第6章的第6节由黑龙江大学冷伏海撰写(现在中国科学院文献情报中心工作);第3章由河北大学顾潇华撰写;第4章第2节有关ISBD和AACR2的内容由东北师范大学徐跃权撰写;第4章中除了有关ISBD和AACR2的内容及信息组织技术标准概述之外的部分由黑龙江省图书馆肖红凌撰写;第5章中除第6节以外的内容由山西大学裴成发撰写;第6章中除第6节以外的内容由黑龙江大学冯璐撰写;第7章由哈尔滨医科大学史继红撰写;第8章由黑龙江大学周燕和安新颖撰写。本书体系框架由冷伏海和史继红设计,由冷伏海统稿。

本书撰写中得到了中国科学院文献情报中心韩平老师自始至终的悉心指教,他不仅给本书的写作提出了许多宝贵意见,而且以科学工作者“严谨细致,勇于奉献”的精神对本书进行了审核,在此表示衷心的感谢!

另外,在撰写本书的过程中参考引用了许多专家学者的成果,在这里深表敬意和谢意!

由于学识限制,时间仓促,书中难免存在不妥之处,还请专家和广大读者批评指正。

冷伏海
2003年3月于北京中关村

目 录

序

前言

1 绪论	(1)
1.1 信息组织的概述	(1)
1.1.1 基本名词术语	(1)
1.1.2 信息组织的必要性	(4)
1.1.3 信息组织的专业含义、要求及目的	(5)
1.2 信息组织的学科发展基础	(6)
1.2.1 信息组织发展的理论基础	(7)
1.2.2 信息组织发展的方法学基础	(8)
1.2.3 信息组织的学科发展研究	(9)
1.3 信息组织的原则	(9)
1.3.1 确立信息组织原则的意义	(9)
1.3.2 信息组织的基本原则	(10)
1.4 信息组织的方法	(12)
1.4.1 优化选择	(12)
1.4.2 确定标识	(14)
1.4.3 组织排序	(18)
1.4.4 改编重组	(19)
1.5 信息组织的类型	(21)
1.5.1 信息组织的类型划分	(21)
1.5.2 检索工具类型的划分	(21)
2 信息组织的发展史略	(25)
2.1 信息组织发展的历史实践	(25)
2.1.1 国内信息组织实践	(25)
2.1.2 国外信息组织实践	(30)
2.2 信息组织的发展与发展趋势	(34)
2.2.1 数字信息环境下信息组织的变化	(35)

2.2.2	数字网络环境下传统信息组织方式的发展趋势	(37)
2.2.3	网络环境下对信息资源新的组织方式的开发	(38)
3	信息组织的环境	(40)
3.1	信息组织的社会环境	(40)
3.1.1	信息社会对信息组织的影响	(40)
3.1.2	知识经济对信息组织的依托	(42)
3.1.3	数字化技术带来信息组织的变革	(43)
3.1.4	信息组织的网络环境	(44)
3.2	信息组织的专业环境	(45)
3.2.1	图书情报机构	(45)
3.2.2	档案馆	(47)
3.2.3	博物馆和艺术馆	(48)
3.2.4	办公室	(49)
4	信息组织的技术标准	(52)
4.1	信息组织技术标准概述	(52)
4.1.1	信息组织技术标准及意义	(52)
4.1.2	信息组织的技术标准类型	(54)
4.2	信息组织技术标准	(56)
4.2.1	著录标准	(56)
4.2.2	机读目录	(69)
4.2.3	置标语言	(77)
4.2.4	文件类型定义(DTD)	(92)
4.2.5	沃里克框架(WF)	(108)
4.3	网络协议	(115)
4.3.1	ISO/OSI 参考模型	(115)
4.3.2	几种主要的网络协议	(116)
4.3.3	图书馆信息搜索交换标准接口	(122)
5	信息组织的技术方法——分类法	(127)
5.1	分类法概述	(127)
5.1.1	分类法的含义	(127)
5.1.2	信息资源分类法及其类型	(128)
5.1.3	科学分类与信息资源分类	(131)
5.2	分类法结构	(132)
5.2.1	类目体系	(132)
5.2.2	标记符号	(136)

5.2.3 说明与注释	(138)
5.3 类目体系的建立	(140)
5.3.1 类目体系建立的原则	(141)
5.3.2 类目分析	(142)
5.3.3 类目关系	(144)
5.3.4 总论与专论类目的区分	(145)
5.4 国内外分类法介绍	(146)
5.4.1 《杜威十进分类法》	(146)
5.4.2 《国际十进分类法》	(148)
5.4.3 《美国国会图书馆图书分类法》	(150)
5.4.4 《中国图书馆分类法》	(152)
5.5 分类与分类检索工具	(153)
5.5.1 分类与分类标引	(153)
5.5.2 分类排架与分类检索工具(或系统)	(158)
5.6 分类自动化的研究与进展	(167)
5.6.1 自动分类标引方法研究	(167)
5.6.2 自动分类系统研究	(169)
6 信息组织的技术方法——主题法	(174)
6.1 主题法概述	(174)
6.1.1 主题法与分类法	(174)
6.1.2 主题法的类型划分	(175)
6.2 叙词语言的词汇控制	(179)
6.2.1 词类和词组控制	(179)
6.2.2 词形和词义控制	(183)
6.2.3 词间关系控制	(188)
6.3 叙词表的结构	(194)
6.3.1 字顺展开	(195)
6.3.2 系统展开	(198)
6.3.3 图形展开	(205)
6.4 国内外主题词表介绍	(206)
6.4.1 《美国国会图书馆标题表》	(206)
6.4.2 《医学标题表》	(208)
6.4.3 《汉语主题词表》	(210)
6.4.4 《中国分类主题词表》	(212)
6.4.5 《社会科学检索词表》	(213)

6.5	主题标引和主题检索工具	(214)
6.5.1	主题标引方式	(214)
6.5.2	主题标引方法	(215)
6.5.3	主题标引基本规则	(222)
6.5.4	主题检索工具和主题检索系统	(226)
6.6	主题标引自动化的实践与研究	(229)
6.6.1	自然语言标引和检索实践	(229)
6.6.2	主题标引全自动化研究	(236)
7	信息组织的技术方法——元数据法	(242)
7.1	元数据概述	(242)
7.1.1	元数据法的基本概念	(242)
7.1.2	元数据的类型及作用	(244)
7.1.3	元数据的格式	(245)
7.1.4	元数据的建立与通用性要求	(249)
7.1.5	元数据的开发应用	(250)
7.2	都柏林核心元素集(DC)	(251)
7.2.1	DC的元素和限定词	(252)
7.2.2	DC的内容编码规则与设计原则	(259)
7.2.3	DC的句法及实例	(261)
7.2.4	DC的扩展	(264)
7.3	其他元数据法简介	(265)
7.3.1	电子文本编码体系(TEI)	(266)
7.3.2	联邦政府信息查询系统(GILS)	(269)
7.3.3	VRA可视资源核心	(274)
7.3.4	FGDC数字地理空间元数据内容标准	(276)
7.4	元数据法的发展与研究	(276)
7.4.1	元数据的应用发展	(276)
7.4.2	对元数据方法的评价	(279)
7.4.3	中国元数据的应用与研究	(280)
8	信息组织与系统建设	(282)
8.1	信息组织系统概述	(282)
8.1.1	信息组织系统	(282)
8.1.2	信息组织子系统	(283)
8.2	网络环境下信息组织的系统建设	(285)
8.2.1	网络环境下信息组织体系的构建	(285)

8.2.2 网络信息组织的基本原理	(287)
8.2.3 网络信息组织的主要工具	(290)
8.3 搜索引擎的信息组织——Yahoo	(291)
8.3.1 搜索引擎的信息组织方式	(291)
8.3.2 Yahoo 的信息组织方式及评价	(292)
8.3.3 Yahoo 的分类结构	(294)
8.3.4 中文 Yahoo! 数据库的搜索式查询	(296)
8.4 多媒体信息组织与管理技术	(298)
8.4.1 多媒体及其技术与系统	(298)
8.4.2 多媒体信息组织管理技术	(299)
8.4.3 基于 Web 的多媒体信息查询框架	(305)
8.4.4 网上多媒体信息的搜集	(306)
8.4.5 多媒体信息的分类处理	(308)
8.4.6 多媒体信息的查询与检索	(309)
8.4.7 中国路桥多媒体数据库	(310)
8.4.8 多媒体信息组织评价	(312)
8.5 EIS 中的信息组织	(312)
8.5.1 EIS 与信息组织	(312)
8.5.2 EIS 的超文本结构	(314)
8.5.3 EIS 的信息组织技术	(315)
8.5.4 EIS 的超媒体查询方法	(317)
8.5.5 EIS 的信息组织评价	(318)
8.6 数字图书馆的信息组织	(318)
8.6.1 数字图书馆的信息组织对象	(318)
8.6.2 数字图书馆信息组织的目标和原则	(319)
8.6.3 数字图书馆信息组织的数据结构	(321)
8.6.4 数字图书馆信息组织的数据库技术	(324)
8.6.5 数字图书馆的系统结构	(328)
8.6.6 数字图书馆的技术体系结构	(330)
参考文献	(334)

1 绪 论

历史车轮的飞快旋转已把人类引领进 21 世纪的大门,随之而来的是由科学技术的发展和人类社会的进步所掀起的更为势不可挡的信息革命浪潮。今天,信息成了继物质、能量之后的“第三级资源”,成了人类宝贵的物质财富和精神财富。越来越多的不同领域的人们正以“信息就是生产力”的信条作为鞭策自己前进的动力,同时要求自己不断提高对信息的敏锐嗅觉和掌握信息的本领。

1.1 信息组织的概述

1.1.1 基本名词术语

(1) 信息 (information)

信息是一个内涵广泛的名词,可泛指信号、音信、消息和数据等。在信息管理科学中主要指具有特定传播、参考和使用价值的这一内容范畴的信息。在本书中,它是作为图书情报学中的一个基础名词出现的,主要是指知识信息。

(2) 信息资源 (information resources)

本书中所指信息资源,在无需特别指定信息类型的情况下,泛指各类型信息资源。

(3) 网络信息资源

迄今为止,对于“网络信息资源”尚没有统一的定义,类似的名称很多,如“电子信息资源”(electronic information resources)、“因特网信息资源”(internet information resources)、“联机信息”(online information)、“万维网资源”(world wide web resources)等等,不一而足。其定义也多种多样,如有人提出“电子信息资源是以电子数据的形式将文字、图像、声音、动画等多种形式的信息存放在光、磁等非印刷纸质的载体中,并通过网络通信,计算机或终端等方式再现出来的信息资源”。也有人认为,网络信息资源为利用计算机网络的各信息资源的总和。

图书馆界人士相对于现有的图书馆文献资源而言,把计算机网络中存在的信息资源称为“虚拟资源”(virtual resource)。也有的使用“联机资源”(online resource)一词。因为现在的许多联机检索系统实际上已经通过因特网提供服务,所以,两者的区别已经变得不那么重要了,完全可以用以往的联机网络提供的信息

资源的延伸和扩展这个意义去理解。日本的一些图书馆学、情报学研究人员则比较倾向于使用“网络信息资源”这一用语。在欧美,网络信息资源(networked information resource)或网络资源的使用(network resource)也相当频繁。

事实上,“网络信息资源”、“电子信息资源”、“虚拟资源”和“联机资源”在本质上并没有根本的区别。但是,“网络信息资源”要比其他的概念更直观、更准确地表达了本书中所讨论的概念。

(4) 信息组织(information organization)

本书中的信息组织是作为信息管理科学体系之下的图书情报学中的一门专业学科所言的,其基本概念是指在系统科学理论指导下依靠专门的技术方法和手段对信息资源进行选择和整理,从而达到使这类资源得以被充分利用的目的。需要指出的是:这里的“选择”不同于传统图书馆学中的“文献采访”、“文献采购”的概念;这里的“整理”也不同于传统图书馆学中的“编目”,可以理解为是“编目”在 20 世纪 90 年代信息资源及其利用技术发展背景下内涵的延伸。

(5) 因特网(internet)与万维网(world wide web)

“因特网”主要指利用网络的物理结构,包括连接所有用户的客户机、服务器和线路。而“万维网”则主要指利用因特网可以访问的所有网址和信息。在本书中为叙述简便,无特别所指的情况下,皆以“因特网”涵盖两者。

(6) 数字图书馆

“数字图书馆”一词由英文 Digital Library 翻译而来,是用数字技术处理和存储各种图文并茂文献的图书馆,实质上它是一种多媒体制作的分布式信息系统,把各种不同载体、不同地理位置的信息资源用数字技术存储,以跨越区域面向对象的网络查询和传播的一个大型信息系统。

(7) 文献(literature, document)

文献是指用一定的技术手段把一定的信息内容记录到一定载体上形成的存储型传递媒体。国家标准定义为:“记录知识的一切载体。按加工记载的目的通常分为一次文献、二次文献、三次文献;按载体材料通常分为印刷型文献、缩微型文献、声像型文献,机读型文献等。文献是信息资源的一种存在形态,在网络环境下,文献类型的范围已经拓宽了,而且两者的界限也逐渐模糊了。”本书主张用第一种定义。

(8) 文献情报机构(document and information organization)

无特指时泛指进行信息组织和提供信息服务之机构。

(9) 著录与标引(标准定义)

著录(description),即“文献著录”或“目录著录”,是对表示文献内容、外表形式和物质形态的特征进行分析、选择和记录的过程,著录的结果是款目。

标引(indexing)指在分析文献内容或情报提问的基础上,用某种索引语言或标识符号把文献的主题概念及其他有检索意义的特征标识出来,作为情报存储和检

索的依据的处理过程。按照使用的索引语言或标识符号的类型,可分为分类标引、主题标引和非主题标引。按照使用标引设备,可分为手工标引、自动标引和机助标引。

(10) 款目与记录(标准定义)

款目(entry)依据一定的方法,对表示文献内容、外表形式和物质形态的特征所做出的记录,即著录的记录,通过记录文献的各种特征,它可以提供一种文献的书目信息,是一种文献的缩影。借助款目可以认识文献,分析比较文献。它是组成目录的最小单位。

(11) 数据与数据处理

数据(data)是能够输入到计算机由计算机处理的数字、字母和符号。

数据处理(data processing)是用某种方法记录信息并通过机械化方法即时获得信息的过程,方式有常规的和倒置的两种。常规的方式,指在一个记录上记载一个文件的各个项目的代码;倒置的方式,则是以一种特征如一个索引词或一个分类号作单位记在一个记录上。

(12) 检索(retrieval)

通常是指从任何方式组成的信息集合中,查找特定用户在特定时间和条件下所需信息的方法与过程。完整的信息检索含义还包括信息的存储。

(13) 数字化信息(digitalization information)与数字信息环境(digital information environment)

数字化信息是网络传播的内容。它可以囊括有史以来人类在文化、艺术和科学领域所取得的一切优秀成果,借助于以计算机信息加工处理为核心的数字化技术,将各种作品中的原始信息,例如语言、图像、文字、符号、数据、表格等全部变成二进制编码,形成可以在网上流动的信息资源。

张晓林教授认为,在数字信息规模发展的推动下,各种影响信息传送和利用的技术、经济、法律和运营机制问题不断得到解决,数字信息环境中的信息内容和服务功能也在迅速建立和完善,基于数字化的信息环境已初步形成。

在此环境中,人们利用 Internet、Intranet(内部网)、VPN(Virtual Private Network,虚拟内部网络)、WAP(Wireless Access Protocol,无线传输网络)、Wireless Internet(无线因特网)、Digital Appliances(数字家电)等方式,在任何时间和地点通过诸如电子邮件、网站、协作系统、虚拟社区等信息交流技术和系统,接入各种数字信息系统来方便地检索和利用数字形态的文件、图书、期刊、图像、音乐等信息内容,并基于信息网络来进行相关的业务运作和交易,从而进入数字化生存状态。

数字化信息与数字信息环境具有以下特征:

1) 任意层次的信息内容元素、信息单元和信息集合体系正逐步以计算机可识别和理解的方式定义、描述、指向、链接、传递和动态组织。