



高等学校
信息管理类专业
核心课教材

信息组织

(第二版)

主编 周宁

武汉大学出版社

高等学校信息管理类专业核心课教材

信息组织

(第二版)

主编 周 宁
编著 周 宁 黄晓梅 文燕平
严亚兰 刘 玮 杨 峰

武汉大学出版社

图书在版编目(CIP)数据

信息组织/周宁主编. —2版. —武汉:武汉大学出版社, 2004. 11
高等学校信息管理类专业核心课教材
ISBN 7-307-04374-2

I. 信… II. 周… III. 信息管理—高等学校—教材 IV. G203

中国版本图书馆 CIP 数据核字(2004)第 106709 号

责任编辑: 严红 陈朋 责任校对: 黄添生 版式设计: 支笛

出版发行: 武汉大学出版社 (430072 武昌 珞珈山)

(电子邮件: wdp4@whu.edu.cn 网址: www.wdp.whu.edu.cn)

印刷: 武汉大学出版社印刷总厂

开本: 880×1230 1/32 印张: 15.25 字数: 418千字 插页: 1

版次: 2001年11月第1版 2004年11月第2版

2004年11月第2版第1次印刷

ISBN 7-307-04374-2/G·699 定价: 25.00元

版权所有, 不得翻印; 所购我社教材, 如有缺页、倒页、脱页等质量问题, 请与当地图书供应部门联系调换。

内 容 提 要

本书是信息管理专业核心课程的教材之一。它系统地阐述了信息资源描述与组织的理论方法和实现技术；重点讨论了网络环境下数字资源的组织。

信息资源的组织方法一般分三级：网络搜索引擎是一级组织方法。它利用知识结构分类和主题描述语言实现，其方法简便易行，但其描述深度有限，因而检索效率不尽如人意。用元数据方案描述和组织信息资源为第二级组织方法。由于其成本低、效率高，受到人们的青睐。采用复杂格式：如利用 MARC（ISO 2709）和 MARC（XML）建立数据库是第三级组织方法。它以描述深度好，检索效率高而备受用户的称赞。

本书专门讨论了网络环境下的多媒体信息组织，广泛采用 XML 技术，对数字图书馆、电子商务、电子政务中的信息组织进行了专题讨论，并对信息组织的高级形式——知识组织进行了探讨。

前 言

“信息组织”是信息管理类专业的核心课程之一，它系统地讲述信息资源描述与组织的理论与方法。随着因特网的迅猛发展，数字化信息资源已成为信息资源的主体。在信息生产与消费的过程中，人们对信息组织的理论、方法、技术等方面都有所创新，并开辟了许多全新的应用领域。因此，本教材着眼于知识经济和信息社会发展的需要，重点讨论以计算机为基础的数字化信息资源的描述与组织；吸收国际上该领域最新成果的内容，全面反映 21 世纪该领域研究前沿的学术进展和主要成果，力争在体系结构、理论方法上有所突破。

第 1 版出版后，经过三年的教学实践，课程内容有了较大的扩充和更新。在此基础上，对第 2 版的结构作了调整，内容有较大扩充。和第 1 版相比，第 2 版增写了三章：

第 5 章 下一代因特网的基础语言——XML；

第 8 章 搜索引擎；

第 11 章 电子政务的信息组织。

根据该学科的科学研究和应用的发展，重写了三章：数字图书馆的信息组织、电子商务中的信息组织和知识组织。另外，对第 1~4 章和第 6~7 章作了较大的修改和补充。

第 2 版全书共 12 章，其主要内容如下：

第 1 章为引论，讲述了信息资源在知识经济中的地位与作用、信息组织的发展沿革、主要内容及常用方法。

信息组织

第2~3章分别论述了信息资源组织的分类描述语言和主题描述语言。它系统地总结了国内外的主要成果和常用方法,并以网络搜索引擎为例,具体讨论了分类描述语言和主题描述语言在因特网中的应用。

第4章讨论了数字信息资源描述与组织的重要标准 SGML 和 HTML 的理论与具体应用、网上信息的组织与发布、多媒体信息的组织方法等。

第5章对下一代因特网的基础语言——XML作了系统论述,结合实践对其功能特点和应用进行了归纳。

第6章介绍了信息资源描述与组织的元数据方法。简述了信息资源的三级组织方法,重点论述 Dublin Core 元数据研究的系列成果、最新定义、描述方法和具体应用,并结合 RDF 等元数据对 RDF/XML 框架的应用作了进一步的研究。

第7章论述了信息资源组织的数据库方法及相关标准,对 MARC21 (2709)、MARC21 (XML) 的理论方法和应用实例作了系统论述。信息组织的数据库方法经历了多种体系结构的发展和应用,对于新近出现的研究成果——半结构化数据库系统也作了专题讨论。

第8章对网络信息资源组织与利用的工具——搜索引擎作了系统论述。从搜索引擎的体系结构、建立与使用方法入手,结合国内外有代表性的搜索引擎作了具体研究,并对元搜索引擎作了探讨。

第9~11章论述了数字图书馆、电子商务系统和电子政务中的信息组织。在理论与实践相结合的基础上,对当今世界的这些热点课题的研究和实践进行了探讨。

第12章论述了信息组织的高级形式——知识组织。讨论了信息组织与知识组织的关系、知识组织的常用方法;对显性知识与隐性知识的转化、知识组织与知识管理的目标进行了研究;重点讨论了图书馆的知识组织和企业的知识组织;同时,还对知识组织工具进行了探讨。

本教材中列举了大量的实例，力求理论与实际紧密结合；还列出了3个实验课题供学员上机实践时使用。

本书按教学大纲要求，由集体编写而成；黄晓梅编写第1~3章，严亚兰编写第9章，刘玮编写第10章，杨峰编写第11章，文燕平编写第12章，其余几章由周宁编写，并负责全书的统稿。

本书的出版得到了武汉大学各级领导的大力支持，武汉大学出版社的同志为此付出了辛勤劳动。在编写中得到了国内一些同行专家的关心、支持和帮助，参考了国内外许多学者的研究成果和论著，在此一并表示感谢。由于信息组织课程涉及的知识面宽、内容广，加之发展快、技术新，虽然我们经过了艰苦努力，但错漏之处难免，欢迎专家、读者批评指正。

编 者

2004年7月 于珞珈山

目 录

1 信息组织引论	1
1.1 信息资源在知识经济中的地位与作用	1
1.1.1 信息在我国经济和社会发展中的作用	1
1.1.2 信息组织在信息资源开发中的地位与作用	4
1.2 信息资源的描述与发现机制	5
1.2.1 信息组织原理	5
1.2.2 信息资源描述	9
1.2.3 信息资源检索	16
1.3 信息组织的发展沿革	18
1.3.1 按职能划分信息组织发展阶段	18
1.3.2 按组织对象划分信息组织发展阶段	21
1.4 信息组织方法与技术	23
1.4.1 信息组织的基本方法	23
1.4.2 基于传统印刷型载体信息的组织方法	25
1.4.3 网络环境下信息资源的组织方法	25
1.4.4 现代信息组织技术	29
2 信息资源组织的分类描述语言	34
2.1 信息资源管理的分类语言	34
2.1.1 分类语言的构成原理	35
2.1.2 分类语言的体系结构	38

信息组织

2.2 国内外分类语言的实例	44
2.2.1 《中国图书馆分类法》	44
2.2.2 《杜威十进分类法》(DDC)	52
2.2.3 《国际十进分类法》(UDC)	55
2.2.4 《国会图书馆图书分类法》(LCC)	59
2.2.5 《冒号分类法》(CC)	60
2.3 数字化信息资源管理中的分类描述语言	61
2.3.1 数字化信息资源组织的特点	62
2.3.2 网络信息组织的分类法	64
2.3.3 网络信息资源分类体系的建立	71
2.4 网站分类实例	73
2.4.1 Yahoo!	73
2.4.2 Canadian Information By Subject	77
3 信息资源组织的主题描述语言	82
3.1 主题描述语言的沿革	82
3.1.1 标题词描述语言	83
3.1.2 叙词描述语言	90
3.1.3 关键词描述语言	98
3.2 国内外主题语言应用实例	105
3.2.1 《汉语主题词表》	105
3.2.2 MeSH	111
3.2.3 INSPEC Thesaurus	113
3.3 网络信息资源组织中主题语言的应用	117
3.3.1 关键词法在搜索引擎中的应用	117
3.3.2 现有主题词表在网络信息资源组织中的应用	119
3.4 信息组织与资源共享的实例	121
3.4.1 OCLC	121
3.4.2 新华网	130
3.4.3 中国价格信息网	133

4 SGML 与 HTML	138
4.1 SGML	138
4.1.1 SGML 简介	138
4.1.2 SGML 的工作原理	140
4.1.3 SGML 的应用	141
4.2 HTML	142
4.2.1 HTML 简介	142
4.2.2 HTML 标签和属性	144
4.2.3 字符样式	154
4.2.4 表格标签和属性	159
4.2.5 表单	168
4.2.6 链接的标记与属性	174
4.3 Web 的信息组织与发布	180
4.3.1 网络信息组织原理	180
4.3.2 Web 信息组织与发布	182
4.4 多媒体信息的组织	184
4.4.1 Web 上的图像信息组织	185
4.4.2 Web 上的声音信息组织	190
4.4.3 Web 上的视频信息组织	193
5 下一代因特网的基础语言——XML	198
5.1 XML 的诞生与发展	198
5.2 XML 文档结构	202
5.2.1 XML 声明	203
5.2.2 元素类型声明	204
5.2.3 属性表声明	205
5.2.4 实体声明	207
5.3 XML 数据建模方法	207
5.3.1 定义 DTD	207
5.3.2 定义外部 DTD	209

信息组织

5.3.3 XML Schema	211
5.4 扩展样式语言	213
5.5 X Link 与 X Pointer	215
5.6 XML 的应用	217
6 元数据	221
6.1 元数据简介	221
6.1.1 元数据的发展	221
6.1.2 元数据的级别	223
6.2 DC	224
6.2.1 都柏林核心元数据系列研究成果	226
6.2.2 都柏林核心元素定义	244
6.2.3 DC 的应用	248
6.3 RDF	252
6.4 其他元数据	255
6.4.1 PICS	255
6.4.2 CDF	256
6.4.3 MCF	257
7 MARC21 与信息组织的数据库方法	260
7.1 MARC21 (2709)	260
7.1.1 MARC21	261
7.1.2 ISO 2709	268
7.1.3 GB 2901 格式	271
7.2 MARC21 (XML)	271
7.3 信息组织的数据库方法	276
7.3.1 单用户数据库系统的信息组织方法	277
7.3.2 主从式数据库系统的信息组织方法	281
7.3.3 分布式数据库系统的信息组织方法	284
7.3.4 C/S 数据库系统的信息组织方法	284

7.3.5 B/S 数据库系统的信息组织方法	285
7.3.6 数据模型的扩展	285
7.4 半结构化数据库系统	287
7.4.1 数据集成子系统	287
7.4.2 数据发布子系统	288
7.5 多媒体数据库	289
7.5.1 多媒体数据库的关键技术	289
7.5.2 多媒体数据库的设计、建立与利用	291
8 搜索引擎	292
8.1 网络搜索引擎的诞生与发展	292
8.2 搜索引擎的系统结构	293
8.2.1 搜索器	294
8.2.2 索引器	294
8.2.3 索引数据库	295
8.2.4 检索器	297
8.2.5 用户接口	297
8.3 搜索引擎的使用方法	297
8.3.1 布尔逻辑检索	297
8.3.2 截词检索	298
8.3.3 词组与名称检索	298
8.3.4 范畴检索	299
8.4 搜索引擎的实例	300
8.4.1 Google	300
8.4.2 百度	307
8.5 元搜索引擎	313
8.5.1 元搜索网站	313
8.5.2 客户元搜索程序	317
9 数字图书馆的信息组织	321

信息组织

9.1 数字图书馆	321
9.1.1 数字图书馆概述	321
9.1.2 数字图书馆发展	323
9.1.3 数字图书馆系统结构	329
9.1.4 基于元数据的信息组织	332
9.2 DC 图书馆应用纲要	333
9.2.1 概述	333
9.2.2 DC-Lib 的体例结构	334
9.2.3 DC-Lib 的基本内容	335
9.3 OAI-PMH	338
9.3.1 OAI-PMH 的相关概念	338
9.3.2 OAI-PMH 的特征分析	343
9.3.3 OAI-PMH 的请求与响应命令	344
9.3.4 DC 核心的 XML 方案	348
9.4 管理元数据	349
9.4.1 管理构件的概念及其组成	349
9.4.2 一个基于 XML 的实例	351
10 电子商务中的信息组织	354
10.1 电子商务简介	354
10.1.1 电子商务的产生与发展	354
10.1.2 电子商务的含义	357
10.2 电子商务的主要模式	359
10.2.1 主要的电子商务模式	359
10.2.2 一个 B2C 解决方案	363
10.3 电子商务系统的体系结构	365
10.3.1 电子商务框架	365
10.3.2 电子商务系统的框架结构	366
10.3.3 电子商务系统模型	369
10.3.4 电子商务概念模型的基本要素	370

10.3.5 实例	371
10.4 电子商务系统的信息组织	375
10.4.1 电子商务系统的设计步骤	375
10.4.2 电子商务网站设计的基本原则	379
10.4.3 购物网站的信息组织要素	380
10.4.4 个性化电子商务网站的设计	382
10.4.5 电子商务网站的建设方法	383
10.5 电子商务的技术标准及安全协议	385
10.5.1 技术标准	385
10.5.2 安全协议	386
10.6 商务网站的实例分析	388
10.6.1 易趣——交易的乐趣	388
10.6.2 亚马逊——网上书店的创始者	401
11 电子政务的信息组织	407
11.1 电子政务的诞生和发展	407
11.1.1 电子政务的产生	407
11.1.2 电子政务的定义	408
11.1.3 电子政务的发展阶段	410
11.1.4 电子政务的发展	411
11.2 电子政务的系统结构与信息组织方法	412
11.2.1 电子政务的系统结构	412
11.2.2 电子政务的信息	416
11.2.3 电子政务的信息组织	417
11.3 电子政务的主要功能及其特点	423
11.3.1 从政府服务的对象看电子政务的功能	423
11.3.2 从电子政务与其他信息系统的比较看其功能	424
11.3.3 电子政务的特点	424
11.4 电子政务的相关的标准与技术规范	426
11.4.1 国际电子政务标准化概况	426

信息组织

11.4.2 我国电子政务标准	427
11.5 电子政务系统的实例	428
12 知识组织	432
12.1 从信息组织到知识组织	433
12.1.1 资料、信息与知识	433
12.1.2 知识组织的提出与发展	434
12.2 图书馆的知识组织	436
12.2.1 图书馆知识组织对象	436
12.2.2 图书馆知识组织技术	437
12.2.3 图书馆知识组织的方式	440
12.3 企业的知识组织	442
12.3.1 企业知识组织对象	442
12.3.2 企业知识组织技术	443
12.4 知识组织的工具	448
12.4.1 知识仓库	448
12.4.2 Topic Map	452
12.4.3 专家系统	454
12.4.4 基于本体的知识组织	455
索 引	458
主要参考文献	465

1

信息组织引论

1.1 信息资源在知识经济中的地位与作用 ■

从茫茫宇宙到微电子，无处不充满着信息（information）。信息是物质存在的一种方式、形态或运动状态，是标记在通讯系统中的信号，是构成知识的原料。《牛津英语词典》综述了信息的三种用法：其一是作为过程的信息（information-as-process）；其二是作为知识的信息（information-as-knowledge）；其三是作为事物的信息（information-as-thing）。信息与物资、能源已成为现代社会的三大支柱。它是继物资、能源之后的“第三级资源”。人们每时每刻都在生产信息、消费信息。从长期的社会实践中人们认识到，信息是人类的宝贵财富，是国家的战略资源。

1.1.1 信息在我国经济和社会发展中的作用 ●

我国第十个五年计划纲要指出：“加强信息资源开发，强化信息资源共享，推动信息技术在国民经济和社会发展各领域的广泛应用”是全国人民在2001~2005年的奋斗目标之一。在当今知识经济时代，信息资源开发与利用的水平已成为衡量一个国家发展状况和综合国力的重要尺度。

长期以来，国内外专家学者和有关的国际组织对信息资源开发、信息社会的形成和知识经济的发展作了大量研究，已取得不少

信息组织

成果。下面引述部分论述，以便我们对其发展历程有一个系统的了解。

1962年，美国经济学家弗里茨·马克卢普，根据美国从第二次世界大战以来至50年代末的社会发展和产业结构变化背景，提出了“知识产业”（knowledge industry）的概念。概念外延包括：①教育；②R&D（研究与开发）；③传播业；④信息设备；⑤信息服务。

1973年，哈佛大学社会学家丹尼尔·贝尔出版《后工业社会的来临》一书，影响广泛，引起激烈的争论。早在1959年夏，在奥地利萨尔茨堡学术讨论会上，他已首次使用“后工业社会”一词；随后，逐步形成“后工业社会”的概念。他定义说：“前工业社会依靠原始的劳动力并从自然界提取初级资源”；“工业社会是围绕生产和机器这个轴心并为了制造商品而组织起来的”；“后工业社会是围绕着知识组织起来的，其目的在于进行社会管理和指导革新与变革；这反过来又产生新的社会关系和新的结构”。由于“后工业社会”的提法含糊，1979年，贝尔承认“信息社会”的概念较“后工业社会”更为确切。

1980年，未来学家阿尔温·托夫勒发表代表作《第三次浪潮》，他说：“我相信我们已处在一个新的综合时代的边缘。”他提出“超工业社会”的概念。“三次浪潮”的划分，渊源于贝尔。托夫勒在其新著《权力转移》里预言：随着西方社会进入信息时代，社会的主宰力量将由金钱转向知识。同年，法国名记者和作家让-雅克·塞尔旺-施赖贝尔也发表《世界面临挑战》一书。作者认为，信息是当今世界最重要而又取之不尽的资源，而自然资源与能源在地球上却日趋枯竭。他直接运用“信息社会”概念取代“后工业社会”和其他提法。

1982年，未来学家约翰·奈斯比特发表《大趋势》一书，从10个方面论述了美国社会发展趋势。他认为：“知识是我们经济社会的驱动力”，“信息经济社会是真实的存在，是创造、生产和分配信息的经济社会。”他概括了信息社会的四个特征，其中包括：