

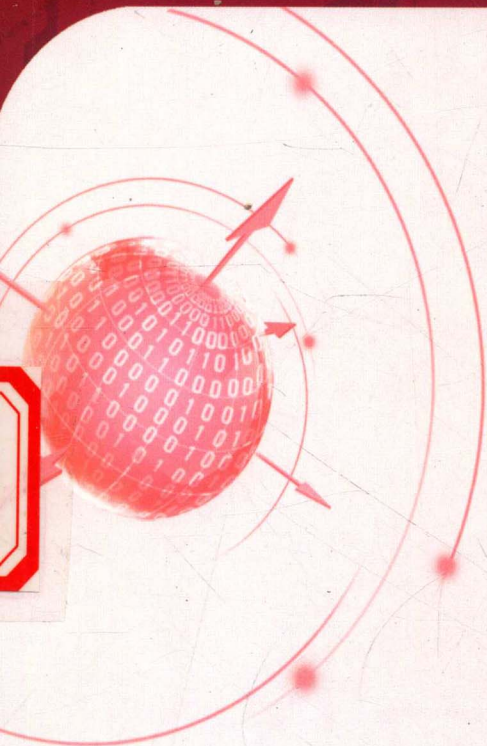


高等学校信息管理专业系列教材

信息组织

第四版

主 编 周 宁
副主编 余肖生 吴佳鑫



WUHAN UNIVERSITY PRESS

武汉大学出版社

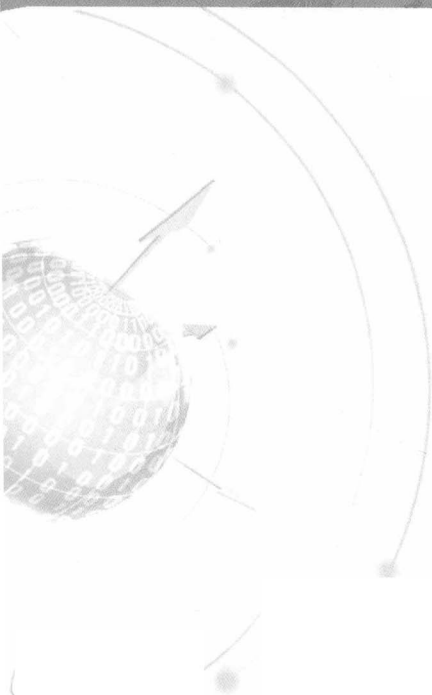


高等学校信息管理专业系列教材

信息组织

第四版

主 编 周 宁
副主编 余肖生 吴佳鑫



WUHAN UNIVERSITY PRESS

武汉大学出版社

图书在版编目(CIP)数据

信息组织/周宁主编. —4版. —武汉: 武汉大学出版社, 2017. 9
高等学校信息管理专业系列教材
ISBN 978-7-307-19668-1

I. 信… II. 周… III. 信息组织—高等学校—教材 IV. G254

中国版本图书馆 CIP 数据核字(2017)第 220737 号

责任编辑: 詹 蜜

责任校对: 李孟潇

版式设计: 马 佳

出版发行: **武汉大学出版社** (430072 武昌 珞珈山)

(电子邮件: cbs22@whu.edu.cn 网址: www.wdp.whu.edu.cn)

印刷: 湖北省荆州市今印印务有限公司

开本: 720 × 1000 1/16 印张: 31.75 字数: 571 千字 插页: 1

版次: 2001 年 11 月第 1 版 2004 年 11 月第 2 版

2010 年 4 月第 3 版 2017 年 9 月第 4 版

2017 年 9 月第 4 版第 1 次印刷

ISBN 978-7-307-19668-1 定价: 58.00 元

版权所有, 不得翻印; 凡购我社的图书, 如有质量问题, 请与当地图书销售部门联系调换。

前 一 版 言 言

当前，我们正处于“互联网+”时代，互联网企业如雨后春笋茁壮成长，中国互联网科技公司三架马车百度、阿里巴巴和腾讯（BAT）在2016年又取得骄人成绩。各行各业都连在互联网上，开发和利用互联网。信息产业的高歌猛进，大力促进了经济发展和社会进步，信息组织出现了一些新理论、新方法、新成果。本教材第4版是在第3版的基础上进行了重写与调整，并新增了如下内容：

1. 增写第10章（原第10章后移为第11章）——社交媒体的信息组织。用一章的篇幅论述了社交媒体的类型与特点，其生成模式和信息组织方法，并以微博和微信为例讨论了其信息组织的具体方法。

2. 在移动互联网环境下，信息组织的语言工具发展迅速。因而在第4章新增了HTML5等内容。

3. 在信息组织的数据库方法中，关系数据库发展较早，使用SQL操作为典型规范。大量的多媒体数据很难用关系数据库组织，因而在第6章增加了NoSQL数据库的内容。

4. 信息资源的计算机化管理沿用了传统的分类法和主题法。为了提高信息组织的质量、发挥计算机管理的优势，在重编第2、3章后新写了本体与本体库、分类主题一体化等内容。

5. 在大数据/云计算环境下，移动互联网迅速发展、智能化异军突起，“大智移云”时代初见端倪，因而在第1章和后面的有关章节中增写了相关内容。

6. 知识组织的理论方法和实现技术发展很快，改写的第11章增加了知识地图和知识图谱等相关的内容。

另外，对第5、7、8、9章中的部分内容进行了改写与补充。

本教材第4版编委会由文燕平、严亚兰、刘玮、杨峰、张会平、余肖生、张芳芳、吴佳鑫、马续补、周宁组成。经过半年多的努力，完成了撰稿与修订工作。

随着大数据、智能化、移动互联网和云计算的快速发展，人类将迎来“大智移云”新时代。信息组织面临许多新的机遇与挑战。

由于信息产业飞速发展，涌现了许多新理论、新技术、新方法、新成果，跟踪研究与应用确属不易。我们虽付出了很大努力，但仍力不从心、挂一漏万。书中错漏之处难免，欢迎读者批评指正。

编者

2017年2月

本书在编写过程中，参考了国内外许多文献资料，特别是近年来出版的有关信息组织方面的书籍和文章，在此一并致谢。本书在编写过程中，得到了许多专家和同行的帮助，特别是北京邮电大学图书馆的同志，为本书的编写提供了大量的资料和数据，在此一并致谢。本书在编写过程中，得到了许多专家和同行的帮助，特别是北京邮电大学图书馆的同志，为本书的编写提供了大量的资料和数据，在此一并致谢。

第一版前言

随着社会信息化,信息社会化程度的不断提高,人类正从工业社会向信息社会迈进。信息是物资、能源之后的“第三级资源”。它是人类的宝贵财富。信息产业是朝阳产业。信息生产与消费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,而信息生产的核心是信息的描述与组织。《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和方法。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,反映 21 世纪的时代特征,力争在体系结构、理论方法上有所突破。

本书共 11 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。

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,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 4~5 章讨论了数字信息资源描述的相关标准(SGML、HTML 与 XML 等)和网上信息(包括多媒体信息)的组织与发布,并简述了信息资源的三级组织方法。

第 6 章介绍了信息资源描述与组织的元数据方法。本章重点论述 Dublin Core 元数据研究的系列成果和描述方法。

第 7~8 章论述了信息资源组织的数据库方法及相关标准。ISO 2709、

MARC、GB2901 都是重要的国际标准和国家标准。在 30 多年的发展中,数据库方法经历了多种体系结构和组织方法。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 9~10 章论述了数字图书馆和电子商务系统中的信息组织。

第 11 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。同时,还对数据仓库和数据挖掘等新课题进行了探讨。

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 4 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,刘玮编写第 9 章,严亚兰编写第 10 章,文燕平编写第 11 章,其余 5 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注、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,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

2001 年 9 月 于珞珈山

第二版前言

《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，全面体现 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。

第一版出版后，经过三年的教学实践，课程内容有了大的扩充和更新。在此基础上，第二版的结构作了调整，内容有较大扩容。和第一版相比，第二版增写了三章：

第 5 章 新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。

第二版全书共 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。

第 2~3 章论述了信息资源组织的分类描述语言和主题描述语言，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,多媒体信息信息的组织方法等。

第5章对新一代因特网的基础语言——XML作了系统论述,对其功能特点和应用结合实践进行了归纳。

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。本章重点论述 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 等元数据与 RDF/XML 框架的应用做了进一步的研究。

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做了系统论述。信息组织的数据库方法经历了多种体系结构的发展和运用。并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,从理论与实践的结合上进行了研究。当今世界的这些热点课题吸引着人们去研究和实践。

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

第三版前言

光阴似箭，信息组织教材第一版出版转眼已过9年。在21世纪的这9年中，信息组织的新理论、新方法和新技术发展很快。本教材虽在2004年第二版中作了重要的更新与补充，但仍不能全面体现信息组织学的新思想和新成果。为此，本教材在第二版的基础上对一些重要内容进行了重写与调整。它主要体现在以下几个方面：

1. 信息组织向集成化、可视化、智能化发展中的一系列新成果。
2. 以新视角重新总结、归纳和补充了信息组织的理论方法、技术基础和研究的主要内容。
3. 由于Web2.0的发展，重写了网络信息组织内容，引入了一些新理念、新成果，补充了搜索引擎与索引数据库的新技术。
4. 考虑到有些院校已先后开设了网页设计与网站建设等课程，所以对第二版的第9、10、11章在重写时合为第9章。浓缩版或许更符合当前的实际。
5. 根据当前发展状况，对元数据内容的选择和利用有新的成果展现和描述。

第三版共10章，具体内容如下：

第1章为信息组织引论，论述了信息组织的诞生、发展历程和在信息产业中的地位与作用，重点讨论了信息组织的理论方法、技术基础和研究的主要内容。

第2章讨论了信息资源管理的分类语言。从它的发展成果总结入手，重点论述了分类语言的原理与体系结构，以及数字化、网络化信息资源的分类方法。

第3章为信息资源组织的主题描述语言。回顾了主题语言的诞生与发展成果，重点论述了它的构成及广泛应用。

第4章系统论述了置标语言工具，对SGML、HTML和XML语言的原

理、相关标准与具体应用进行了讨论，并结合网上信息的组织与发布从理论与实践的结合上进行了探讨。

第5章为信息资源描述与组织的元数据方案，对元数据的发展状况、元数据的结构与分类进行了研究。重点论述了都柏林核心、MARC、METS、MODS等元数据方案的构成及应用，并就元数据间的互操作进行了探讨。

第6章为信息组织的数据库方法。数据库技术是信息组织的最佳技术。首先，讨论了数据库的特点、体系结构和分类方法；然后，具体论述了数据库技术在信息组织中的广泛应用。

第7章讨论了网络信息资源的组织。由于信息资源越来越丰富、Web1.0平台到Web2.0的变化，网络信息资源组织与发布方法也随之变迁。面向用户、面向协作、面向社会、面向机器理解等是当今网络信息组织的新特点。

第8章为索引数据库与搜索引擎。从索引机制入手，系统论述了搜索引擎的诞生与发展，它的系统结构与使用方法，并对元搜索引擎的功能与特点进行了具体探讨。

第9章讨论了信息组织在各个行业的广泛应用，并以数字图书馆、电子商务系统、电子政务系统、数字博物馆为例进行了系统论述。

第10章探讨了信息组织集成化、可视化、智能化的发展趋势，并对信息构建、信息可视化、知识组织进行了专题研究。

信息组织教材与课程建设息息相关、相辅相成。在学校领导的关心和支持下，信息组织学课程获2006年湖北省精品课。2009年，“信息组织学课程建设与创新”获湖北省高等学校教学成果奖。同时，教学课件、课程创新等均获得相关单位的奖励。这是《信息组织》教材的读者长期支持的结果，也是对课程组的鞭策与鼓励。

本教材是集体研究和实践的成果。教材编写人员有：周宁、吴佳鑫、文燕平、严亚兰、刘玮、杨峰、张会平、余肖生、张芳芳、黄晓梅，由周宁、吴佳鑫统稿。

长期以来，本教材的编写得到了许多同行专家的支持和帮助。我们参阅了许多学者的文献，在此一并致以诚挚的谢忱。由于信息组织学涉及的知识既广且深，其新理论、新方法、新技术、新成果不断涌现；要跟踪研究与创新发展，其任务艰巨、任重道远。虽然我们付出了艰苦的努力，但错漏之处难免，欢迎读者不吝指正。

编者

2009年9月

38	1
45	1
45	1
49	4
55	8
54	8
56	9
56	10
57	13
58	13
62	17
63	22
67	22
69	25
102	28
102	30
104	32
100	33
100	34
111	34
111	35

目 录

1	引论	1
1.1	信息与信息组织	1
1.1.1	信息与信息社会	1
1.1.2	信息组织的诞生与发展	4
1.2	信息组织的理论方法	8
1.2.1	信息组织的理论基础	8
1.2.2	信息组织的方法基础	9
1.2.3	信息组织的基本方法	10
1.3	信息组织的技术基础	13
1.3.1	信息组织的四大基本技术	13
1.3.2	信息组织中高新技术的应用	17
1.4	信息组织学研究的主要内容	22
1.4.1	信息资源描述	22
1.4.2	信息资源标引	25
1.4.3	信息资源的整序与存储	28
1.4.4	信息资源检索与利用	30
1.5	信息组织在信息产业中的地位与作用	32
	习题一	33
2	信息资源组织的分类语言	34
2.1	分类语言的原理与体系结构	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
2.2	国内外分类法实例	42
2.2.1	《中国图书馆分类法》	42
2.2.2	《杜威十进制分类法》(DDC)	49
2.2.3	《国际十进分类法》(UDC)	52
2.2.4	《国会图书馆分类法》(LCC)	54
2.2.5	《冒号分类法》(CC)	56
2.3	网络信息资源组织中的分类法	57
2.3.1	数字化信息资源组织的特点	57
2.3.2	网络信息组织的分类法	59
2.3.3	网络信息资源分类体系的建立	64
2.4	分类法应用实例与实践	66
2.4.1	传统分类法的应用	66
2.4.2	自编分类法的实践	71
	习题二	72
3	信息资源组织的主题语言	74
3.1	主题法的原理与体系结构	74
3.1.1	标题词描述语言	74
3.1.2	叙词描述语言	80
3.1.3	关键词描述语言	87
3.2	国内外主题法实例	93
3.2.1	《汉语主题词表》	93
3.2.2	MeSH	97
3.2.3	INSPEC Thesaurus	99
3.3	网络信息资源组织中的主题法	102
3.3.1	关键词法在网络信息组织中的广泛应用	102
3.3.2	现有主题词表在网络信息资源组织中的应用	104
3.4	信息组织方法的发展与应用	106
3.4.1	本体和本体库	106
3.4.2	Folksonomy	109
3.5	分类主题一体化的实例	111
3.5.1	《中国分类主题词表》发展历程	111

3.5.2 《中国分类主题词表》Web 版功能	112
习题三	117
4 置标语言工具	118
4.1 概述	118
4.2 SGML	120
4.2.1 SGML 简介	120
4.2.2 SGML 原理	121
4.2.3 SGML 应用	122
4.3 HTML	123
4.3.1 HTML 简介	123
4.3.2 基本标签与属性	125
4.3.3 字符样式	133
4.3.4 表格标签与属性	137
4.3.5 链接标记与属性	145
4.3.6 多媒体页面	156
4.3.7 框架	161
4.3.8 表单	163
4.4 XML	170
4.4.1 XML 简介	170
4.4.2 XML 文档结构	175
4.4.3 DTD	183
4.4.4 XML Schema	190
4.4.5 CSS	198
4.4.6 XSL	200
4.5 HTML5 及其应用	204
4.5.1 HTML5 新增结构元素	205
4.5.2 HTML5 表单	210
4.5.3 HTML5 音频与视频	216
4.5.4 HTML5 画布与 SVG	218
4.6 JSON	220
4.6.1 JSON 语法规则	220
4.6.2 JSON 取值	220
4.6.3 JSON 对象	221

4.6.4	JSON 数组	221
4.6.5	JSON 应用实例	222
	习题四	222
5	信息资源描述的元数据方案	224
5.1	元数据概述	224
5.1.1	元数据的由来	224
5.1.2	元数据的定义	225
5.1.3	元数据的功能	225
5.2	元数据的分类	226
5.2.1	按元数据的功能分类	226
5.2.2	按元数据复杂程度分类	227
5.2.3	按描述的资源类型分类	228
5.3	元数据的结构	229
5.4	都柏林核心元数据	230
5.4.1	DC 概述	230
5.4.2	都柏林核心元素定义	231
5.4.3	DC 的应用	235
5.5	复杂元数据方案 MARC	239
5.5.1	MARC 概述	239
5.5.2	MARC21(2709)	240
5.5.3	MARC21(XML)	245
5.5.4	ISO 2709 与 GB 2901	249
5.5.5	CNMARC	252
5.6	其他元数据	256
5.6.1	METS	256
5.6.2	MODS	257
5.6.3	EAD	259
5.6.4	GILS	261
5.6.5	PICS	263
5.6.6	CDF	264
5.6.7	MCF	265
5.7	元数据互操作	265
5.7.1	元数据互操作的背景	265

5.7.2	元数据互操作的层次	266
5.7.3	元数据互操作协议	270
	习题五	275
	附录：都柏林核心元数据系列研究成果	276
6	信息组织的数据库方法	296
6.1	数据库组织方法概述	296
6.1.1	数据库组织方式的特点	296
6.1.2	信息资源数据库的分类	298
6.2	数据库系统结构及其信息组织方法	302
6.2.1	单用户数据库系统的信息组织方法	302
6.2.2	主从式数据库系统的信息组织方法	305
6.2.3	分布式数据库系统的信息组织方法	308
6.2.4	C/S 数据库系统的信息组织方法	308
6.2.5	B/S 数据库系统的信息组织方法	308
6.2.6	数据模型的扩展	309
6.3	半结构化数据库系统	310
6.3.1	数据集成子系统	310
6.3.2	数据发布子系统	311
6.4	多媒体数据库	311
6.4.1	多媒体数据库的关键技术	312
6.4.2	多媒体数据库的设计、建立与利用	313
6.5	NoSQL 数据库及其广泛应用	315
6.5.1	关系型数据库与 NoSQL 数据库比较	316
6.5.2	MongoDB	317
6.5.3	Redis	325
6.5.4	Hbase	328
	习题六	334
7	网络环境下信息资源的组织	335
7.1	网络环境的变迁	335
7.1.1	因特网发展的三个阶段	335
7.1.2	Web1.0 与 Web2.0	336
7.2	网络信息资源概述	337

7.2.1	网络信息资源的定义	337
7.2.2	网络信息资源的特点	338
7.2.3	网络信息资源的种类	339
7.3	Web 信息组织与发布	341
7.3.1	Web 信息组织的技术条件	342
7.3.2	网页设计与信息发布	342
7.4	网络环境下信息组织的层次	344
7.5	不同层次的信息组织	346
7.5.1	信息源层的信息组织	346
7.5.2	信息集合层的信息组织	346
7.5.3	信息视图层的信息组织	348
7.6	Web2.0 环境下信息资源的组织	348
7.6.1	面向用户	348
7.6.2	面向协作	349
7.6.3	社会化特征	351
7.6.4	面向机器理解	351
	习题七	358
8	索引数据库与搜索引擎	359
8.1	索引数据库与索引机制	359
8.2	搜索引擎的诞生与发展	360
8.2.1	搜索引擎的诞生	360
8.2.2	搜索引擎的分类	362
8.2.3	搜索引擎的发展趋势	363
8.3	搜索引擎的系统结构	364
8.3.1	搜索器	364
8.3.2	索引器	367
8.3.3	索引数据库	367
8.3.4	检索器	369
8.3.5	用户接口	370
8.4	搜索引擎的使用方法	371
8.4.1	布尔检索	371
8.4.2	截词检索	373
8.4.3	词组与名称检索	373