

普
华
经
管

「信息组织」

01010101010010101110001010011111101001010011000
0101111100001010010101011111111000000010010
0001111010101010101011011001111111001010011100
0011111101010101010101010111111

THE ORGANIZATION OF INFORMATION

曾伟忠 / 编著

 人民邮电出版社
POSTS & TELECOM PRESS

信息组织

曾伟忠 编著

人民邮电出版社
北京

图书在版编目 (C I P) 数据

信息组织 / 曾伟忠编著. -- 北京 : 人民邮电出版社, 2013.4
ISBN 978-7-115-31228-0

I. ①信… II. ①曾… III. ①信息管理 IV.
①G203

中国版本图书馆CIP数据核字 (2013) 第044971号

内 容 提 要

本书共 16 章。首先从理论上对信息组织的历史发展和国内外信息组织的脉络进行了梳理; 然后从图书馆、档案馆、企业和政府的实践工作出发, 探讨了在新的信息环境下的文献分类法、编目法和主题法以及 ERP 信息编码、公文信息组织、电子商务网站信息组织、电子政务网站信息组织和企业信息组织; 最后介绍了网络信息组织中的元数据、HIML 和本体编辑工具 Protégé。

本书可作为大专院校本科生、研究生信息管理与信息系统、图书馆学、情报学、档案学、行政管理等专业的教材或教学参考书, 也可供信息工作者、图书情报档案工作者、企事业相关工作人员学习参考, 还可作为图书馆、情报所、档案馆分类编目专业人员的工具书。

信息组织

-
- ◆ 编 著 曾伟忠
责任编辑 张亚捷
 - ◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街 14 号
邮编 100061 电子邮件 315@ptpress.com.cn
网址 <http://www.ptpress.com.cn>
北京隆昌伟业印刷有限公司印刷
 - ◆ 开本: 787×1092 1/16
印张: 23 2013 年 4 月第 1 版
字数: 500 千字 2013 年 4 月北京第 1 次印刷

ISBN 978-7-115-31228-0

定价: 49.00 元

读者服务热线: (010) 67129879 印装质量热线: (010) 67129223

反盗版热线: (010) 67171154

广告经营许可证: 京崇工商广字第 0021 号

前 言

本书的编写始于作者申报的 2010 年度江西省省级教改课题——基于开放存取模式的信息组织课程教学资源建设和共享研究。

信息组织这门课程是信息管理与信息系统专业、图书馆学专业、情报学专业、档案学专业的核心课程，据统计目前全国有 160 多所院校开设了信息管理与信息系统这个专业。因为信息组织这门课程的前身是图书情报专业的图书分类和文献编目，所以国内信息组织教材的内容大多是文献分类和编目。在教学中作者发现，个别教材偏重于理论，教材中缺乏实践流程，更缺乏企业信息组织和政务信息组织方面的内容，而现在本科生毕业后进图书馆工作的不多，这样培养出来的学生与实际工作要求还有一段距离。基于以上想法，作者有了想编撰一本既重图书馆和档案馆实践工作又有企业信息组织和政务信息组织内容的教材，正好省级教改课题的立项提供了这个契机。

作者以前在图书馆从事过多年的图书分类和编目工作，积累了一些实际工作经验；自 2009 年南开大学图书馆学专业博士毕业后，就一直在南昌大学从事信息管理与信息系统专业本科生和情报学专业研究生的信息组织教学工作。在教学中，针对学生提出的意见，作者在这本教材中除了增加了 ERP 信息编码等企业信息组织及政务信息分类与公文主题标引等内容以外，还增加了电子商务网站和电子政务网站信息组织的内容，这些内容对学生未来的职业发展有很大益处。

本书具有以下几个特点。

一是实用性强。本书的编写目的是培养学生的信息组织实际工作能力，因此在内容

安排上,本书摒弃空洞的理论和概念的罗列,代之以实际工作中需要懂得的技术和知识。通过本书的学习,学生不仅能掌握图书馆和档案馆中外文图书及档案的计算机分类与编目工作,而且能掌握企业中的供应商、客户和物料编码工作以及政务信息的分类和主题标引工作。

二是与其他同类书相比,本书涵盖了档案信息分类和编目的内容,这样便于实现图书情报档案工作的一体化。

三是本书还配有系列教学软件,软件包括中图法分类软件及具有简单编目功能和MARC编目功能的编目软件,需要的读者可通过邮箱 zwzzwz9211@163.com 与作者联系。

四是本书还附有CALIS分类标引及主题标引的标准和规则,可以作为图书馆专业分类编目人员的工具书。

鉴于企业信息组织和政务信息组织两部分是作者前瞻性的探索内容,一定程度上会存在不完善的问题,如果本书存在哪些不足之处,敬请广大读者批评指正。

最后,由衷感谢南昌大学给予本书的出版资助。

曾伟忠

2012年10月于南昌大学前湖

目 录

第一章 信息组织概论.....	1
第一节 信息社会.....	1
第二节 信息组织.....	2
第三节 信息组织的历史发展.....	4
第四节 信息组织的技术.....	6
第五节 信息组织语言.....	7
第二章 分类法与分类工作.....	11
第一节 分类法原理.....	11
第二节 信息分类和编码的基本原则与方法.....	13
第三节 中国图书馆分类法.....	18
第四节 文献分类工作.....	31
第三章 主题法.....	39
第一节 主题法概述.....	39
第二节 中文文献主题标引方法.....	43
第三节 各类文献主题标引方法.....	50
第四节 CALIS 联合目录 CNMARC 字段一览表.....	61
第四章 著录法.....	85
第一节 著录法概述.....	85
第二节 各类文献的著录方法.....	87

第五章 国内外主要分类法和主题法介绍	115
第一节 中国科学院图书馆图书分类法	115
第二节 杜威十进分类法	115
第三节 国际十进分类法	117
第四节 美国国会图书馆图书分类法	119
第五节 冒号分类法	120
第六节 《汉语主题词表》	121
第七节 《中国分类主题词表》	125
第八节 美国国会图书馆标题表	128
第九节 《医学标题表》	130
第六章 西文著录	133
第一节 西文著录项目及说明	133
第二节 各类西文文献的著录方法	135
第七章 中文和西文著录规范控制	147
第一节 规范控制概论	147
第二节 中文名称规范控制	157
第三节 西文著录规范控制	166
第八章 档案分类标引和主题标引	185
第一节 档案分类标引	185
第二节 高等学校档案实体分类法	190
第三节 档案主题标引	195
第四节 革命历史档案主题标引	197
第五节 档案标引质量管理	199
第九章 档案著录与立卷	201
第一节 档案著录标识符和格式	201
第二节 档案著录各项说明	203
第三节 MARC AMC 与 EAD 档案著录标准	210
第四节 档案立卷	213

第十章 ERP 系统与信息编码	219
第一节 ERP 概述.....	219
第二节 ERP 系统各功能模块.....	223
第三节 ERP 数据整理.....	229
第四节 ERP 信息编码概论.....	232
第五节 ERP 部门、客户和供应商编码.....	236
第十一章 ERP 物料编码	241
第一节 ERP 物料编码简介.....	241
第二节 ERP 物料编码方法.....	243
第三节 ERP 物料编码技巧.....	246
第四节 企业 ERP 物料编码实例.....	248
第五节 ERP 信息编码要注意的问题.....	268
第六节 物料编码的修订.....	270
第十二章 政务信息分类	271
第一节 政务信息分类概述.....	271
第二节 政务信息分类具体内容.....	273
第十三章 公文主题标引	289
第一节 公文主题标引规则和程序.....	289
第二节 公文主题标引注意事项.....	293
第十四章 网络信息组织	299
第一节 网络信息分类法.....	299
第二节 自由分类法.....	301
第三节 元数据.....	304
第四节 HTML.....	307
第五节 本体信息组织.....	311
第十五章 电子商务网站信息组织和电子政务网站信息组织	315
第一节 电子商务网站信息组织.....	315

第二节	电子政务网站信息组织.....	324
第十六章	企业网站信息组织.....	335
第一节	企业网站.....	335
第二节	企业网站信息组织.....	337
第三节	企业网站信息组织存在的主要问题及解决方案.....	344
附 录	349
参考文献	359

第一章 信息组织概论

第一节 信息社会

1963年,日本社会学家梅棹忠夫在《朝日放送》杂志上发表了一篇题为《论信息产业》的论文,文中首次提出了“信息社会”这一概念。

信息社会与传统的农业社会和工业社会有着显著的不同。农业社会依靠原始的劳动力并从自然界获取初级资源,土地资源成为当时最重要的资源,劳动为财富之父,土地为财富之母。工业社会是围绕生产和机器这个轴心并为制造商品而组织起来的,资本、技术、劳动力和企业家才能成为重要的资源,谁拥有产品和市场,谁就具有竞争力。信息社会则是围绕着信息运行起来的,其目的在于进行社会管理和革新,信息成为获取一切的重要手段。

信息社会是人类社会发展的一个新阶段,其具有以下几个特征:信息科技高度发达且高度普及;产业结构发生巨大变化;信息和知识成为财富和力量;信息推动人类生活的质量不断提高,社会可持续发展;信息化和全球化是信息社会的两大鲜明特征。

信息化的过程是一个渐变的过程,它首先表露出来并为人们注意到的是“信息产业化”,接着是“产业信息化”,这其中伴随着“产品信息化”、“企业信息化”,之后就是“国民经济信息化”,最后是全社会的信息化。

2000年我国明确提出了“以信息化带动工业化,发挥后发优势,实现跨越式发展”的信息化战略。

信息技术革命为经济全球化提供了必不可少的技术条件,全球化是当代世界经济最根本的特征,信息技术的特点决定了以信息产业为核心的知识经济一开始就是世界性的。随着信息技术革命的不断深入,使不同社会制度和发展水平的国家都不可避免地被纳入一体化的全球体系之中。信息技术的发展使信息的获取变得更为容易,国际贸易和投资也变得更加简单、快捷。与此同时,信息产业自身以及全球贸易、合资合作、跨国联营都取得了长足的进步。信息技术和信息产品的国际交流成为全球知识经济的重要组成部分,包括计算机软硬件和相关服务在内的全球信息技术市场以年均8%的速度增长,无论是发达国家还是新兴的工业化国家和发展中国家,都力争在前途无量的信息市场占有一席之地。

第二节 信息组织

一、信息组织的概念、特征、类型和目的

(一) 信息组织的概念

对于信息组织的概念，国内很多专家都有精彩的论述。

马张华认为：“信息组织亦称为信息资源组织，是根据信息检索的需要，以文本及各种类型的信息资源为对象，通过对其内容特征等进行分析、选择、标引与处理，使其成为有序化的集合。”

柯平认为：“信息组织是信息管理的一个重要环节，它是根据信息的内容特征和外部特征，采用一定的原则和方法，对信息进行加工处理，使之有序可用。”

周宁认为：“信息组织即信息序化或信息整序，也就是利用一定的科学规则和方法，通过对信息外在特征和内容特征予以描述和序化，实现无序信息流向有序信息流的转换，从而保证用户对信息的有效获取和利用及信息的有效流通和组合。”

戴维明认为：“信息组织是为了方便人们检索、获取信息而将庞杂、无序的信息进行系统化和有序化的过程。从广义上来说，信息组织的内容包括信息搜集与选择、信息分析与揭示、信息描述与加工、信息整理与存储。”他还认为，对信息的描述与揭示以及序化是信息组织的中心内容；或者说，信息组织首先是实现对信息的规范控制，包括对信息形式特征的描述和对其主题内容特征的揭示，并在此基础上，采取适当的集成方式予以信息整合，建立有序化的信息系统，实现信息有效检索和利用。可见，信息组织是一个增值的过程。

马费成认为：“信息组织亦称信息整序，是利用一定的规则、方法和技术对信息的外部特征和内容特征进行揭示和描述，并按给定的参数和序列公式排列，使信息从无序集合转换为有序集合。”

笔者认为，信息组织就是利用一定的科学方法和技术，通过对信息外在特征和内容特征进行描述和标引，实现无序信息向有序信息的转换，从而使信息资源达到科学组合以实现增值，促进用户对信息的有效获取和利用。

(二) 信息组织的特征

信息组织具有渗透性、依附性和增值性等特征。渗透性指信息组织存在于各种信息揭示、存贮和检索活动之中。依附性指信息组织无法独立存在，它要以信息的识别、揭示等活动为前提。增值性指信息组织可以增加信息的价值，提高信息传播、检索和利用的效率，是其他信息加工活动和信息利用的保障。

(三) 信息组织的类型

按信息的表现形式，信息组织的类型可分为文字信息组织、图像信息组织、声音信息

组织和视频信息组织；按信息的加工程度，信息组织可分为一次信息组织、二次信息组织和三次信息组织。

（四）信息组织的目的

信息组织的目的可以概括为实现无序信息向有序信息的转换。具体地说，信息组织的目的包括减少社会信息流的混乱程度、提高信息产品的质量和价值、建立信息产品与用户的联系、节省社会信息活动的总成本。

二、信息组织的步骤、原理和工具

（一）信息组织的步骤

信息组织的步骤包括信息选择、信息分析、信息描述和信息存储。

第一步是信息选择，即从采集到的、处于无序状态的信息流中甄别出有用的信息，剔除无用的信息。

第二步是信息分析，即按照一定的逻辑关系从语法、语义和语用上对选择过的信息内容和外在特征进行细化、挖掘、加工整理并归类。

第三步是信息描述，即根据信息组织和检索的需要，对信息资源的主题内容、形式特征、物质形态等进行分析、选择、记录。

第四步是信息存储，即将经过加工整理序化后的信息按照一定的格式和顺序存储在特定的载体中。

（二）信息组织的原理

信息组织的原理主要包括分类组织原理和主题组织原理。

分类组织原理是按照一定知识概念体系的树状等级结构或分面结构来组织信息的主题概念，使信息按学科、专业或范畴、组面等逻辑顺序排列，形成信息资源的逻辑分类体系。

主题组织原理是把用来描述信息资源主题概念的自然语言词汇，通过规范控制作为检索标识，构成简单概念或复杂概念逻辑关系的知识语义体系，再对语词进行字顺排列以达到对信息资源整序的目的。

（三）信息组织的工具

按结构原理，信息组织的工具主要分为分类组织工具和主题组织工具两大类。分类组织工具称为分类法或分类表，又可分为等级列举式分类法、分面组配式分类法及列举—组配式分类法。

主题组织工具称为主题词表，依据主题标识特点和构成原理，又可分为标题表、单元词表、叙词表等。

此外，信息组织工具还有术语词典、规范文档、概念图和语义网络等。

第三节 信息组织的历史发展

一、我国古代信息组织

我国古代信息组织以信息的揭示为基本目的,虽然那时还没有“信息组织”这个名称,信息组织的对象主要是文献,信息揭示的方法主要是对信息的记录和分类,着重于信息外在特征的记录和描述,操作完全由个人以手工方式进行,古代信息组织的代表性成果是目录。我国古代信息组织以文献分类组织为主,形成了七分法、四分法和十二分法等分类体系,我国最大的类书——《永乐大典》就是一部带有主题法性质的检索工具。

七分法以刘向和刘歆的《七略》、王俭的《七志》、阮孝绪的《七录》等为代表;四分法有唐朝魏征的《隋书·经集志》等,该分类法确立了经、史、子、集四部分类系统,清代的《四库全书》总目采用四分法已较为完善;十二分法有郑樵的《通志·艺文略》。我国古代使用最久、影响最大、占主导地位的是四分法。

信息组织在我国已经有二千余年的历史,我国第一部比较成熟的文献分类法始于西汉刘向、刘歆父子编制的《七略》,《七略》是我国第一部综合性图书分类目录。公元前26年汉成帝河平三年,刘向等人校勘当时皇家藏书一万多卷,并为每种书编撰叙录一篇,后将各书叙录汇编成册取名《别录》,分成六部:第一部“六艺”,包括诗、书、礼、易、乐、春秋、御、论语、孝经、小学九类;第二部“诸子”,包括儒、道、阴阳、法、名、纵横、杂、农、小说十类;第三部“诗赋”,包括屈原、陆贾、孙卿、杂赋、歌诗五类;第四部“兵书”,包括兵权谋、兵形势、兵阴阳、兵技巧四类;第五部“数术”,包括天文、历谱、五行、蓍龟、杂占、形法六类;第六部“方技”,包括医经、经方、房中、神仙四类;共计三十八类,六部前有总序一部合为七部。刘向死后其儿子刘歆把《别录》简化,编为《七略》,意思是存《别录》七部之略。全书分为七大类,分别为辑略、六艺略、诸子略、诗赋略、兵书略、数术略和方技略,合称《七略》,其中辑略是写在六略之前的一篇概括性的学术简史。这七部(书目只是六部)分类法是现存有文献可考的我国第一部最有系统的图书目录,对后世有着极大的影响。《别录》和《七略》在唐末已经佚失,现今我们只能从东汉班固以《七略》为蓝本编成的《汉书·艺文志》中窥视《七略》的概貌,《汉书·艺文志》也就成为了我国现存最早的目录学专著。

刘歆父子建立的六部三十八类图书分类法并没能沿用至今,因西汉藏书在东汉献帝之时遭遇劫难,魏文帝曹丕时代“采掇遗亡,藏在秘书中外三阁”,由秘书郎郑默据以编撰成《中经》,形成了后来四部分类法的雏形。《中经》这部目录由于亡佚,难以确知其分类,但据晋武帝泰始十年秘书监荀勖校理群书编成的《中经新簿》(也称《晋中经簿》)可以确知其采用了四部分类法。《中经新簿》类目分为:甲部,有六艺及小学诸书;乙部,有古诸子家、近世子家、兵书、兵家、术数;丙部,有史记、旧事、皇览簿、杂事;丁部,有诗赋、图赞、汲冢书。到了唐代,魏徵、长孙无忌等修撰的《隋书·经籍志》用经、史、子、集的名称和顺序取代了甲、乙、丙、丁四部。此后历代的公私藏书目录均采用这种四部分类

法。因唐朝初年设立的集贤书院是按经、史、子、集分四库储藏图书，名为“四库书”，这便形成了“四库”名称的由来。清代修《四库全书》仍沿用着经、史、子、集的名称和顺序，只是对子目进行了一些调整。由于纂修《四库全书》出自钦定，《四库总目》成书后，四部分类法便占据了当时分类界的统治地位，此后称四部分类法为四库分类法。

《四库全书》的排列体系如下。

经部，包含易类、书类、诗类、礼类、春秋类、孝经类、五经总义类、四书类、乐类、小学类共十大类，其中礼类又分周礼、仪礼、礼记、三礼总义、通礼、杂礼六属，小学类又分训诂、字书、韵书三属。

史部，包含正史类、编年类、纪事本末类、杂史类、别史类、诏令奏议类、传记类、史钞类、载记类、时令类、地理类、职官类、政书类、目录类、史评类共十五大类，其中诏令奏议类又分诏令、奏议两属，传记类又分圣贤、名人、总录、杂录、别录五属，地理类又分宫殿疏、总志、都会郡县、河渠、边防、山川、古迹、杂记、游记、外记十属，职官类又分官制、官箴两属，政书类又分通制、典礼、邦计、军政、法令、考工六属，目录类又分经籍、金石二属。

子部，包含儒家类、兵家类、法家类、农家类、医家类、天文算法类、术数类、艺术类、谱录类、杂家类、类书类、小说家类、释家类、道家类共十四大类，其中天文算法类又分推步、算书两属，术数类又分数学、占候、相宅相墓、占卜、命书相书、阴阳五行、杂技术七属，艺术类又分书画、琴谱、篆刻、杂技四属，谱录类又分器物、食谱、草木鸟兽虫鱼三属，杂家类又分杂学、杂考、杂说、杂品、杂纂、杂编六属，小说家类又分杂事、异闻、琐语三属。

集部，包含楚辞、别集、总集、诗文评、词曲共五大类，其中词曲类又分词集、词选、词话、词谱词韵、南北曲五属。

《四库全书》是我国古代最大的图书集成，总计四部、四十四类、六十六属，收录古籍三千五百零三种、七万九千三百三十七卷，装订成三万六千三百册、六千七百五十二函，共抄七部分藏于北四阁（北京故宫的文渊阁、沈阳故宫的文溯阁、北京圆明园的文源阁、承德避暑山庄的文津阁）和南三阁（镇江金山寺的文宗阁、扬州的文汇阁、杭州的文澜阁）中。《四库全书》基本上包括了社会上流传的各种图书，就著者而言包括妇女、僧人、道家、宦官、军人、帝王、外国人等在内的各类人物著作，而把章回小说、戏剧著作排除在外。

我国古代图书分类法的不足之处在于深受“罢黜百家、独尊儒术”之害，过度重视儒家著作，把儒家著作放在突出的位置；把儒家经典放在四部之首，把诸子著作放在子部之外。此外，轻视自然科学，仅收录了农家、医家和天文算法类少数科技著作，这些缺陷在一定程度上影响了我国科学技术的发展。

二、近代信息组织

近代是信息产品进一步丰富、信息管理逐步开放的时期，体系分类法的确立、主题法的研究和应用、索引文摘的发展打破了信息组织中书目形式一统天下的局面。

在近代,国外有美国图书馆学家、教育家杜威于 1876 年编制了《杜威十进分类法》,奥特勒和拉封丹编制了《国际十进分类法》,美国国会图书馆编制了《美国国会图书馆图书分类法》等。

国内有沈祖荣、胡庆生于 1917 年编制了我国第一部以标记符号代表类目的《仿杜威书目十类法》,刘国均于 1929 年编制了《中国图书馆分类法》及叶圣陶编制了《十三经索引》。

三、现代信息组织

到了现代,信息组织的环境发生了巨大的变化:一是信息资源的数量和质量发生了巨大变化,二是信息组织技术环境发生了巨大变化,三是用户需求发生了巨大变化,四是信息组织工作对象发生了巨大变化,五是信息组织成果共享和国际化发生了巨大变化。

现代信息组织顺应了以上这些变化,吸收了先进的信息组织技术方法,确立了以用户为中心的指导思想,其特征表现为多元化的信息资源组织、检索语言的发展和现代信息技术的应用等,具体包括分类法的改造、主题法的发展、自然语言检索系统的兴起等;其中,自然语言检索系统的兴起表现为自动标引、自动分类技术的探索和应用。在分类主题一体化方面,英国学者艾奇逊于 1969 年成功研制出了世界上第一部分类主题一体化的《分面叙词表》。

我国现代图书分类法的发展经历了从最初 20 世纪 20~30 年代模仿杜威法,到全面探索适用我国的理论和方法的过程,初步形成了具有中国特色的分类理论和分类法。

信息组织的发展趋势是理论多元化、方法集成化、技术智能化、内容揭示深入化、系统标准化和工具易用化。

第四节 信息组织的技术

信息组织的技术有计算机技术、多媒体技术和数据库技术等。

一、计算机技术

传统的信息组织是以文献或以“文字”为处理单位,基于手工操作,速度慢、效率低。计算机技术的出现并应用于信息组织工作,极大地促进了信息组织手段与技术的现代化。1956 年,美国海军兵器试验站首次将计算机引入到书目工作,该站采用 IBM-701 型计算机建立了世界上第一个书目存取系统,同期卢恩首创题内关键词索引,开机编索引之先河。进入 20 世纪 60 年代,书目工作自动化进一步发展,机读目录 MARC 研制成功,极大地促进了信息组织现代化的进程。自 20 世纪 70 年代以来,信息组织的现代化领域不断扩展,计算机自动索引、文摘、检索网络化、翻译自动化、编目自动化等都得到了大发展。自动标引技术也相继问世,特别是有关自然语言的自动标引将信息揭示与标引推向了一个崭新的发展阶段。

总之,计算机技术是信息组织的核心技术,随着计算机技术的不断发展,信息组织将逐渐实现信息处理的自动化及智能化。

二、多媒体技术

多媒体技术是计算机技术、通信技术和大众传播技术等有机统一体，是诸项技术相互融合而成的。多媒体技术使计算机由单纯文字和数字处理进化成综合性、网络化及交互性处理声音、文本、图像、图形、动画、电影、电视等多种媒体的综合信息系统。多媒体技术具有多维化、非线性、集成性、交互性和实时性等特征，多媒体技术改善了用户与计算机之间、用户与用户之间交互关系的方法和手段，引起多媒体系统里信息组织方法的变化，从而产生新的显示效果。

三、数据库技术

数据库是存储在计算机内的、有组织的、可共享的数据集合。20世纪60年代，在文件管理的基础上，人们开发了信息管理的新技术——数据库技术。在长期的实践中，数据库以其独特的优点表明它是信息组织与共享的最佳方法。数据库技术从诞生到发展，已经经历了第一代数据库（层次式数据库与网状数据库）应用和第二代数据库（关系数据库）的辉煌。目前，第三代数据库（面向对象数据库）正处于快速发展时期。面向对象的第三代数据库不仅能管理常规数据，而且能管理绚丽多姿的现实世界中的各种信息，多媒体数据库系统就是一个实例。多媒体数据库为现实世界中各种信息的数字化、规范化管理提供了技术保证。

第五节 信息组织语言

一、信息组织语言的概念

信息组织语言是根据信息组织的需要而开发的人工语言，专门用于各种手工的和计算机化的信息标引和检索系统，是表达信息主题概念和检索课题概念的语言，主要是分类语言和主题语言。信息组织语言是表达一系列概括信息内容的概念及其相互关系的概念标识系统，它可以是从自然语言中精选出来并加以规范化的一套词汇，可以是代表某种分类体系的一套分类号码，也可以是代表某一类事物的某一方面特征的一套代码，用以对信息进行主题标引、逻辑分类或特征描述。

二、信息组织语言的类型

按照不同的划分标准可将信息组织语言分为以下不同的类型。

（一）文献外表特征语言和文献内容特征语言

按构成原理，信息组织语言分为文献外表特征语言和文献内容特征语言。

文献外表特征语言有题名语言、著者语言、号码语言等：题名语言按文献题名字顺逐字排检，西文起首冠词一律不排；著者语言按著者姓名以姓前名后的字顺逐字排检；号码语言按文献代码，如专利号、标准号、报告号、ISBN号、ISSN号等的顺序排检。

文献内容特征语言有分类语言和主题语言：分类语言是用分类号和类名来表达信息的主题概念，并且按照知识门类的逻辑次序将信息系统地组织和划分的语言；主题语言是用主题词或标题词来表达文献主题概念的语言，以主题词表或标题词表作为词汇规范化的工具。

（二）先组式语言和后组式语言

按组配方式，信息组织语言分为先组式语言和后组式语言。

先组式语言指在文献标引之前，表达文献信息内容的标识已经事先组配好了的信息标引语言，也称列举式语言，如分类语言。

后组式语言指在标引文献时，不预先规定组配关系，而是在进行标引时，用户按照组配规则临时组配起来的信息标引语言，如主题语言。

（三）规范化语言和自然语言

按照规范化程度，信息组织语言分为规范化语言和自然语言。

规范化语言又称为受控语言，是指人为地对标引词和检索词的词义进行控制和管理语言。简单地说，规范化语言是一种由主题词表或者分类表控制的检索语言，包括主题语言中的叙词、标题词和分类语言中的分类号。

自然语言是直接从原始信息中抽取出来的未经规范化处理、用以揭示信息主题概念的自由词，如关键词语言，它除一般的事物名称、科学术语外，还包括俗名、商品型号和缩写等。自然语言具有不用编制词表、及时跟上事物发展、准确表达事物新概念、选词灵活方便、专指性强、标引和检索速度快等优点，便于计算机检索。

三、信息组织语言的特点

信息组织语言包括词汇和语法等内容。词汇是分类表、词表、代码表中的全部标识，一个标识（分类号、检索词、代码）就是它的一个语词，而分类表、词表、代码表则是它的词典；语法是指如何创造和运用哪些标识来正确表达文献内容和信息的需要，以有效地实现信息检索的一整套规则，其又分为词法（主要用于分类表、词表、代码表编制过程）和句法（主要用于信息标引和信息检索过程）两部分。

信息组织语言必须符合唯一性、规律性、定型性、通用性、准确性和政治思想上的正确性等质量要求，必须排除自然语言中的多词一义、一词多义和词义含糊现象，并要有适当的专指度。

信息组织语言采用等级结构、参照系统、轮排聚类法、范畴聚类法和图示法等各种显示概念之间关系的方法，来实现对内容相同及相关的信息加以集中或揭示其相关性这项功