



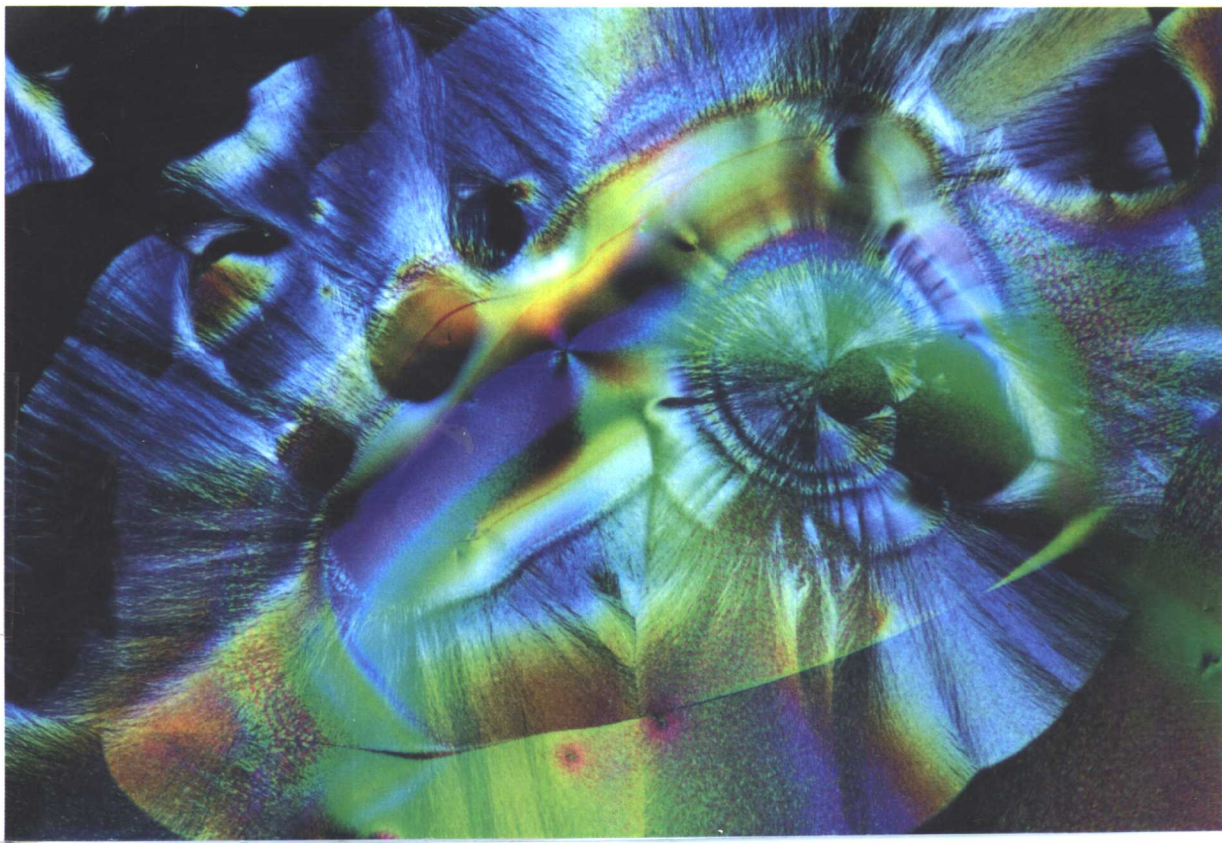
面向21世纪课程教材

信息管理与信息系统专业教材系列

信息组织

马张华 编著

清华大学出版社



面向 21 世纪课程教材
信息管理与信息系统专业教材系列

信 息 组 织

马 张 华 编 著

清华大学出版社

(京)新登字 158 号

内 容 简 介

本书为信息管理专业系列教材之一,系统讲述了信息组织的基本理论方法,包括:信息组织基本知识、信息组织原理、信息描述的理论和方法、分类法理论和国内外典型分类表评介、分类标引与分类检索工具的建立、主题法理论和国内外典型主题词表评介、主题标引与主题检索工具的建立以及自然语言检索与标引等。

本书可作为高等院校信息管理专业、图书馆学、情报学专业教学的教材,或成人教育、函授教育的教材,也可供各种与信息资源组织有关的机构或部门,包括各类型文献机构、网络公司以及企业信息资源管理部门的工作人员学习参考之用,或供对信息资源组织感兴趣的读者阅读使用。

书 名: 信息组织

作 者: 马张华 编著

出 版 者: 清华大学出版社(北京清华大学学研大厦,邮编 100084)

<http://www.tup.tsinghua.edu.cn>

责任编辑: 魏荣桥

印 刷 者: 北京市清华园胶印厂

发 行 者: 新华书店总店北京发行所

开 本: 787×960 1/16 **印张:** 20.25 **字数:** 430 千字

版 次: 2001 年 11 月第 1 版 2001 年 12 月第 2 次印刷

书 号: ISBN 7-302-04650-6/G·213

印 数: 3001~5000

定 价: 28.00 元

序 言

人类正迈向知识经济的新世纪,为了使我国高等教育能适应国内外急剧变化发展的形势,1995年春,原国家教委正式推出和开始实施了“高等教育面向21世纪教学内容和课程体系改革计划(以下简称教学内容改革计划)”。改革的总目标是:“转变教育思想,更新教育观念,改革人才培养模式,实现教学内容、课程体系、教学方法和手段的现代化,形成和建立有中国特色的社会主义高等教育的教学内容与课程体系,提高教学质量,培养适应21世纪需要的社会主义现代化建设者和接班人。”近期目标是:“用5年左右时间,集中优势力量,开展集体攻关,努力形成一批优秀的研究和改革成果,使我国高等教育的教学内容和课程体系相对落后于科技、经济、社会发展的状况有较大改观,为21世纪初叶大范围提高我国高等教育质量打下良好基础。”在高等教育改革的系统工程中,教学改革是核心,而其中教学内容和课程体系的改革,又因其直接反映教育目的和培养目的,是培养人才素质、提高教育质量的核心环节,故又是教学改革的重点和难点,是深层次的教学改革。为了反映和固化各专业基础课和主干课教学内容和课程体系改革的成果,教学内容改革计划还提出要编写出版一批高水平和高质量的“面向21世纪课程教材”。

作为国家教委级221个大的立项项目之一(理科27个之一),“信息管理类专业教学内容和课程体系改革(以下简称信息管理专业教学改革研究)”,也于当年立项并开始启动。项目由北京师范大学牵头,有中国科学技术信息研究所、复旦大学、北京大学、武汉大学、中山大学、四川大学、山东大学、天津师范大学、北京联合大学和吉林工业大学等11所高校和科研院所参加。4年来,经过参加人员和广大教师积极认真的改革研究和实验,取得了重要进展,明显提高了实验班级毕业生培养质量的实际效果,并编写出和推荐了头一批“信息管理与信息系统”专业教材,本书就是这批书中的一种。

以下几点是“信息管理专业教学改革研究”项目考虑教学内容和课程体系改革问题的基本出发点,也是编写这套教材要努力做到的。学生是教学工作的主体,使用这套书的学生,在学习每门课时了解整体学术背景,无论是对加深当前学习内容的理解,还是对认识和发挥自己“主体”地位的作用都有必要。

1. 关于“信息管理”和“信息管理学”

“信息管理”是一个范围很宽,正在发展的概念。按项目组目前研究的认识,“信息管理”主要是指信息资源的管理,包括微观上对信息内容的管理——信息组织、检索、加工、服务等,及宏观上对信息机构和信息系统的管理。“信息管理学”是一门研究人类信息管

理活动的规律及应用的学科,它以数学、管理科学、信息科学与技术作为基础,涉及多学科(理、工、农……)及多领域(经济、管理、法律……)的综合性交叉学科。高附加值的信息是知识创新的关键因素,但浩如烟海的信息,只有经过有效获取、科学加工和有序管理,才能成为可利用的资源,信息管理学科所研究和提供的就是这种理论知识和技术。在国内外经济知识化、社会信息化迅速发展和激烈竞争的形势下,信息管理、信息科学、信息技术、信息产业与信息教育共同成为信息社会战略上需要优先发展的行业。有人把网络比作信息高速公路,计算机软硬件比作在高速公路上跑的“车”,信息资源比作“货”,当前的问题是“车”严重空载和“货”的质量不高,信息管理就是“加工生产”大量优质信息产品的知识工具,而问题的症结就在于需要培养出大批高素质的信息管理人才。

2. “信息管理学”专业的基本内涵

什么是决定“信息管理学”专业核心教学内容和课程体系的基本内涵?应当说目前还没有明确的界定,但根据学科的发展和社会定位,根据信息科技发展对其产生的影响,根据培养目标的定位和国内外许多学者的建议,“信息管理学”的专业内涵应建立起一个由上到下的信息管理学的学科体系。如 Wilson 的观点:“应从信息的产生,信息的组织与处理,信息的传播与利用来定义。”有的学者还提出,一个完整的信息管理科学的知识体系应该包括信息的上、中、下“三游”。而作为一个信息管理和服务的专业人员,应该要“了解信息的上游,掌握信息的中游,提供下游的服务”。

所谓信息的上游,乃是了解信息的本质、信息的特征、信息的种类、信息产生的过程、信息生产的主要机构及信息如何传播等。而这里所谓的信息,不应只限制在学术信息,也不应只强调图书媒体,而应该包括所有相关的信息媒体,及所有可能的相关信息的研究。

所谓信息的中游,则包括信息的搜集、信息的整理、信息的组织与分析、网络环境下多媒体信息的描述与组织、信息系统、信息的储存、信息的检索、信息的深加工等。

而下游的信息利用与服务,则是了解我们的使用者,了解他们的信息需求,然后提供他们所需的信息服务。

如果我们依照信息管理人员的工作内容来看,信息上游的了解是所有信息管理人员所应具备的知识。这个基本常识掌握后,才能因工作性质的不同,调整应掌握的知识面。

信息的组织与分析应是本专业最核心的专业知识,本专业所有其他的专业知识都应该是由此而发展,这也是我们跟别的专业有所区别的地方。

由以上分析是否可认为本专业的内涵是:信息的产生和传播,信息的组织分析,信息的利用与服务。以上认识,也是我们提出分层次的专业基础课和核心课内容体系的基本出发点。

3. 转变观念,建立新的培养模式

为满足对信息管理人才的需要,在新修订的我国高等教育本科专业目录中,在原科技

情报、信息学、经济信息、管理信息系统等专业基础上,调整合并设立了新的“信息管理与信息系统”专业。1999年全国有150多所高校在这个专业招生,成为今年理科最热门的专业之一。如何办好这个新调整的专业,提高培养质量?关键在于转变教育观念,改革培养模式。信息管理是科技含量高、知识更新快、交叉学科多、国际化程度高、应用范围广、服务性强的发展中的新学科,要求本科培养的人才具有广泛和多层次的适应性,为此必须改变以适合在某一种行业或领域从业为培养目标的旧观念,而把拓宽加深基础和以创新能力和实践能力为主的全面素质作为培养目标;改变过去重理论、轻实践、多灌输、少自学的旧教学模式,而把充分开发潜能,使学生做到“自奋其力,自致其知”,作为提高人才素质的运作机制,突破传统基础观念对“加强基础”的理解,不仅要通过系统的理论知识学习打下学生聚合思维的基础,还要通过教学和实践环节的系统改革,培养学生对科学问题的兴趣和锲而不舍的探索精神,打下创造能力所需要的发散思维和人格品质的基础,并要将这两种思维进行整合,内化成学生的认知结构。如开设将信息的采集、描述、组织、存储、检索、分析和利用等若干知识点互相结合应用,充分体现自主学习的“课程设计”课;在教学计划中安排整个第八学期进行面向科研、应用开发和服务等实际问题的毕业实践(毕业设计和毕业论文)等。教学改革是一项系统工程,按系统工程原理,局部和可能大于也可能小于整体,各门课的改革和教材编写要从整体出发考虑,才能实现系统总体目标的优化。

4. 信息科技的发展和教学内容的改革

信息对知识经济发展的意义,可分为两方面:一方面是信息成为知识创新的因素;另一方面,为了有效获取和利用信息,要促使信息科技的发展。特别是90年代以来,网络科技的迅猛发展,推动了知识经济的加速发展。这使得知识经济和信息科技成为大学教育面临的社会背景和重要技术背景,这无疑对信息管理学科内容和体系的变化首当其冲地产生重大影响。因此,信息管理专业教学内容和课程体系改革的重点和难点,不仅要重视内容和体系改革的整体优化,精简陈旧内容,加强信息科技内容的比重,给学生更多实践、研讨和自主学习的时间,积极应用信息技术于教学手段和方法,更重要的是专业教学内容的定量化和信息化,在信息的产生传播、信息的组织分析、信息的利用与服务三个层面的专业教学内容中,以信息科技和相关科技的发展为融合点和生长点,拓宽理论基础,改造传统教学内容,改进研究方法,改变“两层皮”的状况,做到有机结合,形成严谨的专业内涵,构造新的内容和体系。这样才能从根本上开发学生的潜能,培养出面向未来和面向世界的高质量信息管理人才。

5. 信息素养教育

信息科技,特别是网络科技的迅猛发展,也导致人类的沟通与信息交换方式由过去的人际互动(human-human interaction)模式,变为以人机互动(human-machine interaction)

为主的模式。终身学习、如何学习、能力导向学习和开放学习成为新的教育理念的重要内涵。为满足知识创新和终身学习的需要,发达国家纷纷将信息素养(information literacy)教育,作为培养 21 世纪人才能力的重要内容。所谓“信息素养”或“信息能力”,是指使用计算机和信息技术高效获取、正确评价和善于利用信息的能力。目前美国全国从小学、中学到社区学院和各大学都已全面将信息素养纳入正式的课程设置之中。“信息素养”是一个有时代性的、重要的教育议题,是未来信息社会考察一国人力素质和生产力的重要指标。作为信息管理专业的学生,不仅比其他专业更有必要提高自己的信息素养,还有责任在对大众实施信息素养教育计划的重任中发挥作用。

6. 教材特点

首批推荐的这一套教材,反映了经过该项目组成员和广大教师的努力而取得的改革成果。具有新的教育观念和內容,作者水平较高,校际合作、老中青合作范围较大,适应面较广等特点。

本套书的作者都是国内信息管理方面的专家或后起之秀,本套书不仅是他(她)们多年教学改革及经验的总结,也是他(她)们最新科研成果的结晶。有的教材出版印刷过几次,获过奖,此次再版增加了新的研究成果;有几位青年作者近年刚从国外深造归来,在他(她)们的书中反映了学科前沿的最新进展。

经过专业目录调整,全国有多所高校设立了信息管理与信息系统专业,尽管原有专业背景不同,但改革的方向和目标将逐步融合。此外,国家自考大专和自考大学本科也新增加了信息管理与服务专业,不少院校的夜大和函授大学也有此专业,还由于信息管理是发展较快的新兴学科,故可估计此套书出版后的读者群将比较大,尽可能做到兼顾不同层次的读者,也是编写这套教材时重视的一个问题。

一本好的教材一定要经过多次修订才能成熟起来。这套书中有的是第一次正式出版,尽管编著者和出版社为此付出了很大努力,但难免有不足之处,欢迎读者批评指正。

这批教材能列入“面向 21 世纪课程教材”及时出版,要衷心感谢教育部高教司理工处领导的大力支持,也要衷心感谢高等教育出版社、清华大学出版社、武汉大学出版社编辑同志的积极配合与大力支持。

康仲远
于北京师范大学
1999 年 9 月 10 日

目 录

第一章 导言	1
第一节 信息组织的必要性	1
第二节 信息组织的形式和检索途径	3
第三节 信息组织的处理方式和研究内容	7
第二章 信息组织原理	10
第一节 对信息组织的要求	10
第二节 信息组织中的控制与规范	12
第三节 规范控制的依据	18
第三章 信息描述	26
第一节 信息描述的意义	26
第二节 信息资源描述项目、标识和描述级次	28
第三节 检索点和规范文档	39
第四节 计算机编码	45
第四章 信息描述工作	53
第一节 信息描述工作概述	53
第二节 信息描述方法	62
第五章 分类法	74
第一节 分类法概述	74
第二节 分类法结构剖析	78
第三节 类目体系的建立	94
第六章 国内外分类法介绍	111
第一节 《杜威十进分类法》	111
第二节 《国际十进分类法》	116
第三节 《国会图书馆图书分类法》	121
第四节 《中国图书馆图书分类法》	124
第五节 《雅虎》(Yahoo!)	130

第七章 分类标引与分类检索工具	135
第一节 分类标引概述.....	135
第二节 分类标引规则.....	149
第三节 分类索书号与分类检索系统.....	160
第八章 主题法	172
第一节 主题法概述.....	172
第二节 叙词语言的词汇控制.....	177
第三节 叙词表的结构.....	188
第四节 叙词表的编制与维护.....	205
第九章 国内外主题词表介绍	212
第一节 《美国国会标题表》.....	212
第二节 《医学标题表》.....	218
第三节 《汉语主题词表》.....	224
第四节 《中国分类主题词表》.....	230
第五节 《社会科学检索词表》.....	236
第十章 主题标引和主题检索工具	241
第一节 主题标引概述.....	241
第二节 主题标引方法.....	244
第三节 主题标引规则.....	259
第四节 主题检索工具.....	275
第十一章 自然语言标引与检索	287
第一节 自然语言概述.....	287
第二节 关键词语言.....	288
第三节 文本检索.....	292
第四节 后控词表.....	298
第五节 自动标引.....	301
第六节 自然语言与控制语言性能及两者结合的趋势.....	309
主要参考文献	314
后记	315

第一章 导 言

第一节 信息组织的必要性

随着现代科学技术的发展,社会逐渐从信息社会向知识社会迈进,信息资源的开发利用日益受到广泛重视。目前信息资源的特点是数量大、种类多、来源广,特别是计算机网络的逐步普及,数字文本复制的便利和自由发表的可能性,使得信息资源数量急剧增加。但是社会信息量的增长并不意味着用户获取信息的增长,恰恰相反,无序的信息资源不仅无助于信息资源的使用,反而会加剧信息增长与使用的矛盾。人们生动地称这种情况是“信息超载,知识缺乏”。越来越多的人认识到,“原始信息本身并不能产生价值”,只有将其有效组织,按特定的需要集中和揭示,才能产生价值。要有效开发利用信息资源,必须采用相应的方法加以控制和处理,信息组织是关键措施之一。

什么是信息组织?信息组织亦称为信息资源组织,它是根据信息资源检索的需要,以文本及各种类型的信息源为对象,通过对其内容特征等的分析、选择、标引、处理,使其成为有序化集合的活动。

要正确了解信息组织的概念,首先必须了解数据和信息的含义及其区别。所谓数据,一般是指经过直接观察获得的事实,它是无生命的,未经过处理的。确切地说,是“以正式的、适合人类或自动方式交流、转换、处理的对于事实、概念、程序的表述”。信息则是经过人类处理的数据,它是供交流共享的知识,“是指人们通过惯常方式赋予数据的意义”。从这个意义上说,我们日常所说许多信息系统,实际上是数据系统,我们日常所说的信息处理,实际上是指数据处理。因此,在我们的讨论中,应注意信息组织与数据组织的区分。

社会信息作为一种公共交流的知识,是通过不同的形式进行传播的,包括口头方式、出版方式、电子方式等。各种信息媒体都是信息存在的形式,同时也是交流的主要形式。以记录形式进行信息交流,能够克服面对面直接交流在时间上和空间上的局限,使得人类社会可以在继承已有经验的基础上向前发展,对人类社会的发展意义重大,是人类社会进步的重要条件。各种类型的信息资源,既是公共知识的集合,同时也是社会发展的智力资源,是信息组织主要的处理对象。

信息资源一词,在本书中与文献的含义接近,通常是指一切记录形式存在的信息载体,比较而言,只是更侧重于对新型媒体形式的强调。信息资源范围广泛、种类繁多,就信息资源的基本类型而言,它可以是书籍、图书、期刊、报纸、论文、标准、档案等各种类型;就存在的形式而言,它可以是各种传统的印刷型资源,也可以是缩微型、机读型资源以及非

实体资源——如存在在网络上的信息资源。信息组织就是以各种信息媒体形式为对象进行的组织。就其处理的单元而言,信息组织存在着多种不同的层次:它可以直接以信息资源的存在单元为处理对象,例如,以一图书、期刊、网站等为处理单元;也可以以期刊、报刊、网站中的个体,如论文、新闻、网站中的构成部门等为处理单元;还可以直接以其中的信息成分作为处理的单元,进行组织和揭示。

必须指出,信息组织的目的,是建立起信息资源收藏系统和检索工具,方便信息资源的开发利用。因此,要完整了解信息资源的含义,还必须了解它在信息资源开发领域中的地位,特别是弄清信息组织与信息检索的关系。

信息组织是一种为了信息检索的需要对信息资源进行有序化组织的活动,它是与信息检索活动密切联系的。从图 1-1 可以看出,任何对信息检索系统的完整论述,一般必须包括信息组织和检索两个方面。

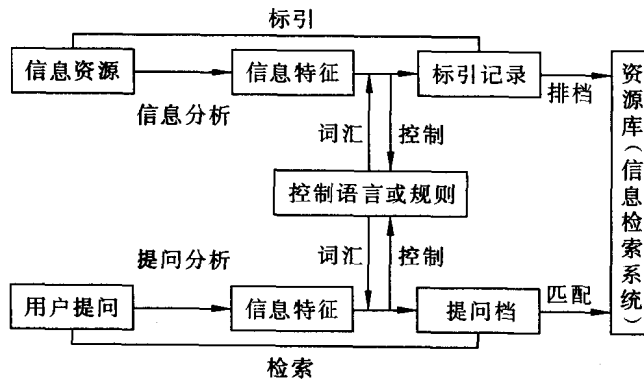


图 1-1 信息整序与信息检索语言

在信息检索系统中,信息组织是依据信息资源的主题内容或特征,将无序的信息资源组织为有序集合的过程。信息组织是信息检索的基础,通常是根据检索的需要,以信息资源的一定单元为单位,记录信息资源的特征,并根据需要对部分信息资源进行必要的相关处理,然后将这些信息资源、记录或条目,按规定的方法输入系统,组成文献资源库或信息检索系统。为了保证处理能达到预定的目的,信息组织中对资源的处理,通常以预先制订的规范为依据,如描述记录资源特征的描述规范、作为资源主题内容标识依据的检索语言及相应的处理规则等。

信息检索则是根据需要,从检索系统中检出相应信息资源的过程。信息检索是信息组织的目的和归宿,是信息组织的反向过程。一般首先依据该检索系统的特点,将检索提问转换成相应检索特征,再根据系统组织方式提供的途径进行匹配查找,检出相应信息资源。在这一过程中,一般应按照系统使用的描述规范及控制语言的特点进行操作。

从信息组织和检索的过程可以看出:信息组织是信息检索的基础和前提,信息资源只

有经过处理和有序化组织后,才可以对它有效地进行查找检索;信息检索是信息组织的出发点和归宿,是信息组织的真正原因。因此信息组织只有根据信息检索的需要实施,才能取得较好的效果。两者是既有区别,又相互依存、相互作用的。这就决定了两者是不能分离的。一方面,信息组织必须充分了解检索需要,根据信息资源的特点和用户的使用要求进行。另一方面,信息检索必须在已有信息组织的基础上进行,离开信息存储和处理的条件,就无法有效地检索。

信息组织和检索中使用的、用以描述检索特征的控制语言及规则,则是根据文献或信息资源组织和揭示的需要制订的组织规范。包括信息资源描述规范,情报检索语言等。前者如文献的编目规则及各种元数据规范等,后者则是一种作为主题内容标引依据的人工语言系统,包括各种分类表或标题表、叙词表以及分类标引和主题标引的规则系统等,一般用于对所使用的词汇及其相关性进行规定,是文献存储和检索的语言保障。信息标识选择和组织的规范是保证信息组织过程中处理的质量有一致性的重要依据,是信息组织和检索的重要工具。也有一些系统对信息资源主题内容的标引不依据词表,但这类系统一般仍需要遵守一定的标引和处理规范。

第二节 信息组织的形式和检索途径

信息组织的目的,是为了便于检索各种类型的信息资源,必要时也同时对文献单位的信息资源进行有序化组织。

信息组织的一个任务是直接组织信息资源本身。即按照一定的方式,将信息资源本身直接纳入信息资源集合。这一组织对传统文本文献的使用尤其重要,是文献单位信息组织的重要内容。

文献组织一般有两种基本排列方式,其一为固定排列法,其二按一定意义的次序排列。固定排列法通常按照文献的形式、到馆先后以及具有固有次序的号码,如专利号、标准号、报告号等排列文献资源;一些不适合按信息资源特征动态组织的资源类型,一般均采用固定排列的方法。这类文献资源包括档案、博物馆的收藏品,此外如光盘、录音带、缩微品等,也大多按照固定排列的方式收藏,通过检索工具加以揭示。

按一定意义的次序排列包括按照题名字顺排列、责任者排列等,但使用最多的是按分类方式排列。由于以分类方式组织文献资源,可以通过内容特征上的联系提供检索的途径,供用户直接对文献资源进行检索,为使用提供方便,因此一般文献单位均采用分类排列的方法。采用这一方法进行文献库组织时,对于同时具有两种以上内容的文献,通常只能按其中一个排列,无法同时按其他内容进行组织;此外,为了方便读者使用,某些文献如工具书、特定对象的文献如儿童读物、盲人读物等往往需要单独组织,这些情况都会造成同一内容文献的分散。至于期刊、报纸中文献的内容,就更无法在文献组织中揭示。因此一般应根据整体情况规划,同时通过检索工具加以揭示。

与信息资源组织不同,检索工具一般以文献记录作为文献的替代物,可以方便地按统一的方式,从多个角度、多种层次组织检索工具和系统。对于检索工具的类型,可以有多种不同的区分方法。按照检索工具的对象和特点,信息组织可以分为以下类型:

- ① 文献目录。通常是以各种媒体的文献单元为对象,对其进行揭示、报道的工具,包括文献收藏目录和书目。前者以一个或多个文献单位收藏的文献为对象加以揭示;后者则以某一领域、地域、或时限内出版的文献为对象进行组织,为用户提供较为全面的文献信息。
- ② 索引。是以文献或文献集中包括的信息内容为其揭示对象的信息检索工具,分为期刊论文索引、报纸索引、工具书索引、语词索引、书后索引等。它一般可以给用户提供某一内容、特征的查找线索,其揭示深入到文献所包含的信息单元,将其与目录结合,就构成了对不同层次文献信息的有效揭示体系。
- ③ 机读数据库。是一种依托现代计算机技术、以机读形式建立的检索系统。数据库通常由某种同类记录按照一定的方式组织而成。按照数据的性质,可以分为文献数据库和非文献数据库两类。文献数据库中,书目数据库通常收录文献目录或索引数据,是手工文摘索引刊物和文献目录的机读形式;全文库则收入文献全文或部分原文数据,用户可以从多种形式入手查阅文献原文。此外,还设有各种非文献库,包括事实数据、数值数据、图像数据等多种类型。机读数据库数据完备,检索手段先进,一般同时提供多种查找方法,并可以通过联机方式查找使用,受到广泛的欢迎。目前许多传统的手工目录、索引正在逐步向机读形式转换,是发展十分迅速的检索工具。
- ④ 网络搜索引擎。这是一种以网络信息资源为对象的检索系统,实际上是数据库的一种特殊形式。与一般的数据库相比,网络搜索引擎涉及的资源类型种类更多,动态性更强,形式更为多样。它不仅可以处理各种类型网站、个人主页等网上资源,而且还可以连接包括聊天室、BBS、网上即时播报等新形式以及各种电子形式的数据库,通过超文本链接的方式,访问各种形式的信息资源。

这样,根据对信息资源中不同对象和层次揭示的需要,上述信息资源检索工具构成了以不同检索需要为目标的、形式多样的、完备的系统。在上述的不同形式中,机读数据库包括目录和索引,同时又增加了新的信息资源组织类型和形式;网络搜索引擎则是对前几种形式的又一次综合和发展,范围更广,包括的资源对象种类更多,可供检索的工具也更丰富。

按照检索工具依据的标识特征和提供的检索途径,信息组织可以分为以下两种类型:

一类为直接依据信息资源的外部特征为标识作为组织依据。常见的这类标识包括:责任者名、题名、机构名、出版地、出版社、标准书号、专利号、档案号、文献登录号等。按这类标识进行组织一般可以通过已知的信息资源特征,得到明确的查找结果。有的特征,如标准书号、专利号,可以据以查出相应的特定信息资源;其他特征,如责任者名、题名等,通常可将范围迅速缩小到便于查检的数量;此外,还可以通过数种因素的组配进行查找,如

通过出版者与出版时间、题名与出版者等的组配,对特定信息资源进行查找。

此外,还有一种根据引证关系建立的检索途径。这种方法将一文献写作过程中引用的文献作为组织检索系统的依据,供用户从引证关系入手对文献进行检索。这一途径也可以通过文献提供的数据直接进行编制。

另一类检索途径则以表达信息资源主题内容的标识为依据。主要有分类法和主题法两种。分类法以表达信息资源内容的标记符号为标识,按照依据主题之间关系进行系统组织,主题法直接以表达文献内容的语词为标识,依据主题字顺的方式提供检索途径。使用主题内容途径检索,不必预先了解信息资源的具体特征,可以直接从所需要的主题内容入手进行查找。例如,在主题检索系统中,可以直接使用“信息组织”一词在检索系统中进行查找,查出论述该主题内容的有关文献。同时还可以根据“信息组织”这一概念与其他主题词的联系,进行相关内容文献的查找和使用。

可以看出,上述两类的检索途径都是十分有价值的,前一种途径可供用户从已知文献数据的角度进行查找,后一类途径则可以根据用户的需要,从需要的主题内容出发对文献进行检索。比较而言,前一种途径可以迅速、准确地查找特定文献,但需要预先掌握确切的数据,否则就无法使用,此外,也很难在此基础上扩大对相关文献的查找;后一种途径则不必事先了解相应的数据,用户不仅可以根据使用需要直接检索特定主题内容的文献,而且还可以通过系统中主题内容之间的联系,扩大或缩小检索范围,进行相关资源的查找,对于信息资源的开发利用,帮助更大,但对特定信息资源查找,不如前一种方法直接和准确,两者是相互补充的。

为了满足用户的不同需要,各种检索工具在对文献信息进行多层次揭示的同时,一般均提供这两类检索途径,方便用户从不同角度检索查找。

按照标识记录的方式,检索系统则可以分为资源单元方式和标识单元方式两种。

资源单元方式以信息资源为中心记录资源,组织检索工具。这类检索系统通常以资源的描述记录代表资源进行组织,作为排检依据的检索标识被记录在该资源的描述数据之中。这种方式形式直观,可以在按标识检索时直接阅读资源的记录,是传统文献单位常用的方式。这种方式的不足是,一个对象在检索系统中只能有一个检索的入口。在手工主题检索工具中,当一个主题涉及多种主题成分时,一般必须组配成一个明确的次序,例:将“河流水污染对鱼类的影响”的标记确定为:河流水污染—影响—鱼类。这一标记在按字顺组织的检索工具中,只能从第一个词“河流水污染”出发才能查到,否则就无法找到。由于这类系统的检索标识,在进行检索以前就已经在检索系统中组配好了,因此,也称为先组式检索系统。

标识单元方式,则是以标识为中心记录资源,一般将一资源的号码记录在相应标识下,所有的标识排列成为一个系统。检索时,先查出相应的标识,再根据标识下的号码查出对应资源,如果涉及复杂主题资源的查找,可以提供相应主题词的组配进行查找。这种方式形式简练,在通过标识的组配进行检索时,不存在检索入口的问题,便

于进行多途径检索。不足之处是,这类系统一般不能进行浏览,比较抽象,在手工系统中,检出的结果只是文献号,还需要逐一查出对应的文献或资源。由于这类系统中检索标识的组配,是根据用户提出检索要求后再进行,因此,这类系统也称为后组式检索系统。

由于两者的特点不同,在实际使用中,两者往往结合使用。例如,在印刷型书目工具中,一般在采用资源单元的同时,结合采用标识单元方式,用户可以根据标识下的文献号,查找对应资源的有关数据;在机械检索系统中,通常以资源为单元完整输入数据,同时提供从标识角度组配查找资源的机制,检出后即调用描述记录,供用户了解资源的描述信息,通过两者的结合,提供检索的便利。

此外,信息组织还可以按照检索工具采用的形式,分为卡片式、书本式、缩微式、电子方式等;按照信息资源对象的类型,分为图书、期刊、报纸、专利、标准、档案、图像、会议文献等;按照检索系统的功用,分为提供文献线索、提供事实、数据、提供信息资源原文等多种形式。表 1-1 为一些常见的信息组织类型。

表 1-1 信息组织的常见类型

信息组织划分标准	信息组织类型
按组织对象分	信息资源
	检索工具
按检索工具的类型分	目录
	索引
	数据库
	网络搜索引擎
按检索对象分	图书检索系统
	论文检索系统
	新闻检索系统
	会议文献检索工具
	专利检索工具
	标准检索系统
	档案检索系统
	博物馆藏品检索系统
	网络资源检索系统
	图像检索系统
.....	

信息组织划分标准	信息组织类型
按照检索途径分	题名检索工具
	责任者检索工具
	分类检索工具
	主题检索工具
	按照引文关系的检索工具

按照组配的方式	先组式检索工具
	后组式检索工具
按照设备	卡片式检索工具
	书本式检索工具
	缩微式检索工具
	电子方式检索工具

第三节 信息组织的处理方式和研究内容

从总体而言,信息组织是根据检索的需要和信息资源的特点,将信息资源组织成有序的系统的过程;就个体而言,则是根据信息组织的整体要求和规范,对信息资源进行描述和处理,并将处理结果纳入系统的过程。一般应根据使用需要,明确对信息资源进行处理的要求。传统文献单位对文献处理的基本目标主要有以下五个方面,包括:

- ① 识别、确认需要处理的所有的物理媒介的文献资源;
- ② 确认文献中或其部分中的著作;
- ③ 按照标准引文规则生成这些文献和著作的系统;
- ④ 提供有价值的检索点,至少包括题名,责任者和主题等;
- ⑤ 在文献单位或其他检索集合中提供某种评论资源的方式。

上述方面同样也是信息组织中对信息资源处理的基本目标。在信息检索工具中,对信息资源的组织,一般是依靠信息记录作为信息资源的代替物实施的。信息记录通常是信息处理的结果,是对各种媒体信息进行多层次、多角度组织和揭示的基础。信息资源的处理,其基本做法是:

- ① 对信息资源进行描述。包括:著录或编写题录,通过对信息资源外部特征进行记录,

用于识别或确认该资源。

- ② 对信息资源的内容进行浓缩。包括编写内容提要或文摘,使用户能迅速了解其内容和特点,供检索查找时选择、判断。
- ③ 提供检索点。主要是确定检索特征及可检信息,以便据此对信息资源进行有序化组织。信息资源著录中描述的各种特征,如:责任者名、题名、出版机构名等,均可用作检索点。同时还可以在对主题内容进行分析的基础上确定相应检索标识,包括分类号和主题词。在计算机检索系统中,文摘和提要中的语词亦可以作为检索依据。

这一信息处理形成的描述结果记录了信息资源的基本数据,是组织检索系统的基本单元。将众多的描述记录按照组织检索工具的需要,以一定的检索点为依据进行排列,就构成了相应资源的检索系统。

从上面可以看出,信息组织活动包括两个基本方面:

其一,信息组织的操作,指根据信息资源的内容和特征,对信息资源进行描述、标引,并将其组织进相应系统的过程。这一操作一般是依据一定的信息资源描述处理的规范,对信息资源的特征和内容进行分析、归纳、选择、记录,并在此基础上将其组织进信息资源集合和检索工具。其中,信息描述通常应与系统处理的要求一致,必须遵循一定的描述或著录的条例;对主题内容的组织和揭示,则需要在对信息资源的主题内容进行分析的基础上,以一定的类表或词表为工具,将分析出来的结果转换为标识,纳入相应的资源库以及检索系统,供用户检索时,进行匹配查找。

其二,信息资源组织规范,是根据信息组织的需要和信息资源的特点预先确定的一系列规则和方法的系统,是进行信息组织操作的依据。包括:

信息描述规范。描述规范是对信息资源特征进行描述记录的依据。通常根据特定信息资源的特点和识别、组织的需要加以规定。一般确定需要记录的项目、记录的文字、次序等,以便可以依据它,以统一、有效的方式对信息资源加以记录。

检索语言。检索语言是一种由表达信息资源主题概念及其相互关系的词汇及规则组成的人工语言系统,国内亦称为情报检索语言。它是根据检索系统的特点,按照使用需要和对知识关系的理解预先组织好的一种信息资源组织方法的体系,是信息组织方法的具体体现。检索语言由词汇和语法两部分组成。其中,词汇是检索语言的主体,主要用于对所使用的词汇及其相关性进行规定,是信息资源存储和检索的语言保障。其表现形式是各种词表,包括分类表或标题表、叙词表等,是根据一定的知识关系,预先建立的一个标识系统或组织标识系统的规范。语法,则是分类标引和主题标引的规则系统,它们规定词汇集的使用方法,是文献标引时必须遵守的规范。自由标引与检索属于自然语言应用,但在一定程度上仍需要遵守一定的语法规范。

此外信息组织规范还应包括对信息资源及检索工具的各种基本组织、检索的方法和规定等。

根据上面的分析,信息组织领域的研究对象一般包括以下基本内容: