

高等院校教学用书

实用文献信息检索

何晓萍 罗时民 主编

实用文献信息检索
实用文献信息检索
实用文献信息检索
实用文献信息检索
实用文献信息检索
实用文献信息检索
实用文献信息检索
实用文献信息检索



机械工业出版社
CHINA MACHINE PRESS



本书共分8章,包括三部分:第一部分(第1、2章)基础篇,概括阐述了信息检索概论、信息检索方法和技术;第二部分(第3、4、5、6、7章)应用篇,系统阐述了期刊文献检索、学位论文文献检索、专利文献检索、会议文献检索、电子图书检索;第三部分(第8章)提高篇,重点阐述了学术文献检索与信息服务。

本书为高等院校教学用书,可供本科院校、高职高专院校学生使用,也可供工程技术人员参考。

主编 罗时民 副主编 何晓萍 参审 姜婷 封面设计 王奕文 责任编辑 王海峰 策划编辑 王海峰 责任印制 洪汉军

图书在版编目(CIP)数据

实用文献信息检索/何晓萍,罗时民主编. —北京:机械工业出版社,2007.9

高等院校教学用书

ISBN 978-7-111-21946-0

I. 实… II. ①何…②罗… III. 情报检索—高等学校—教材
IV. G252.7

中国版本图书馆CIP数据核字(2007)第111052号

机械工业出版社(北京市百万庄大街22号 邮政编码100037)

策划编辑:王海峰 责任编辑:王海峰 王德艳 版式设计:张世琴
责任校对:姜婷 封面设计:王奕文 责任印制:洪汉军

北京铭成印刷有限公司印刷

2007年9月第1版第1次印刷

184mm×260mm·16印张·395千字

0001—8000册

标准书号:ISBN 978-7-111-21946-0

定价:25.00元

凡购本书,如有缺页、倒页、脱页,由本社发行部调换
销售服务热线电话:(010)68326294

购书热线电话:(010)88379639 88379641 88379643

编辑热线电话:(010)88379756

封面无防伪标均为盗版

前言

便捷、高效地获取信息是从事大学教学、科研与社会服务活动的基础。为高水平的人才培养和科学创新研究提供充分的信息保障,是大学图书馆的使命。在这个全球竞争的时代,取得信息固然很重要,但是能够迅速、准确地找到信息、而且能够有效地利用这些信息则显得更为重要。大学生要想在信息时代的激烈竞争中取胜,最优化地完成学业,就必须具备利用现代信息技术获取知识的能力和利用信息解决问题的创新能力。只有这样,才能有效地掌握学习技能,提高学习效率,建立起一个不断演进的知识体系,成为本世纪所需要的高素质复合型人才。

随着数字资源平台的升级和数据库版本的更新,检索模板更趋向于人性化,要求检索内涵更加深入和准确。通过本书的学习能使学生掌握信息检索技术的基本知识和技巧,并充分利用信息资源。本书集不同载体的实用资源和最新检索技术为一体,具有数字化、网络化的时代特征。

本书共分8章,包括三大部分:第一部分(第1、2章)基础篇,概括地阐述了信息检索概论、信息检索方法和技术。第二部分(第3、4、5、6、7章)应用篇,系统地阐述了期刊文献检索、学位论文文献检索、专利文献检索、会议文献检索、电子图书检索。第三部分(第8章)提高篇,重点阐述了学术文献检索与信息服务。

本书由何晓萍教授、罗时民副教授任主编,梁红妮、袁芳、张彧、胡德华参加编写;何晓萍、张彧负责编写写作大纲和全书的统稿工作。本书由博士生导师扶名福教授任主审。

本书编写分工如下:何晓萍(第1章、第2章第1、2、3节、第3章第10、11、12、13、14、15节、第5章第1、2、3、4节、第6章);梁红妮(第7章、第8章第1节);袁芳(第3章第1、2、3、4、5、6、7、8、9节);张彧(第2章第4节);罗时民(第4章、第5章第5节);胡德华(第8章第2节)。

限于编者的学识水平,书中疏漏和不足之处在所难免,敬请学术界同仁和读者批评指正。

编者

目 录

前言	43
第1章 信息检索概论	1
1.1 基本概念	1
1.1.1 基础术语	1
1.1.2 文献类型	1
1.2 信息检索与检索系统	4
1.2.1 信息检索	4
1.2.2 检索系统	4
1.2.3 检索系统的组织	5
1.2.4 检索系统的构成	7
1.3 检索语言	8
1.3.1 检索语言的类型	8
1.3.2 体系分类语言	8
1.3.3 人工语言和自然语言	11
1.3.4 受控语言	12
第2章 信息检索方法和技术	13
2.1 信息检索方法	13
2.1.1 检索途径	13
2.1.2 检索方法	14
2.1.3 检索程序	14
2.2 数字信息检索技术	17
2.2.1 检索方式	18
2.2.2 检索功能	18
2.2.3 检索步骤	19
2.3 网络环境下信息检索	20
2.3.1 信息检索模式	20
2.3.2 网络信息资源组织与揭示	20
2.3.3 网络信息检索策略	21
2.3.4 CALIS、OPAC 检索	21
2.4 网络环境下检索技术	27
2.4.1 Web 搜索引擎	27
2.4.2 文献阅读器	35
2.4.3 文献输出技巧	39
第3章 期刊文献检索	43
3.1 中国期刊全文数据库	43
3.1.1 数据库简介	43
3.1.2 检索指南	43
3.1.3 检索结果处理	49
3.2 万方数据资源系统	51
3.2.1 学位论文全文数据库资源	51
3.2.2 会议论文全文数据库	56
3.2.3 科技信息子系统	56
3.2.4 数字化期刊子系统	58
3.2.5 商务信息子系统	61
3.3 维普科技期刊全文数据库	63
3.3.1 数据库简介	63
3.3.2 检索指南	63
3.3.3 检索结果处理	71
3.3.4 其他功能简介	72
3.4 人大复印报刊资料全文数据库	74
3.4.1 数据库简介	74
3.4.2 检索指南	75
3.4.3 检索结果处理	77
3.4.4 其他功能	78
3.5 Elsevier Science 全文数据库	81
3.5.1 数据库简介	81
3.5.2 检索指南	81
3.5.3 检索结果处理	86
3.6 Springer Link 全文数据库	86
3.6.1 数据库简介	86
3.6.2 检索指南	86
3.6.3 检索结果处理	91
3.7 World SciNet 全文数据库	91
3.7.1 数据库简介	91
3.7.2 检索指南	91
3.7.3 检索结果处理	95
3.8 EBSCO host 系统全文数据库	95
3.8.1 数据库简介	95

3.8.2 检索指南	96	5.1.3 专利权及其特征	169
3.8.3 检索结果处理	101	5.1.4 授予专利权的条件	169
3.9 APS/AIP 数据库	103	5.1.5 不授予专利权的主题	170
3.9.1 数据库简介	103	5.1.6 申请日与优先权	170
3.9.2 检索指南	104	5.2 国际专利分类表	170
3.9.3 检索结果处理	113	5.2.1 概况	170
3.10 Ei Engineering Village 2	114	5.2.2 国际专利分类表的编排结构	171
3.10.1 数据库简介	114	5.2.3 国际专利分类法的分类原则	172
3.10.2 检索指南	115	5.2.4 使用国际专利分类表的辅助 工具书	172
3.11 SCI (Science Citation Index)	122	5.3 德温特专利索引及其检索	175
3.11.1 数据库简介	122	5.3.1 概况与特点	175
3.11.2 检索指南	123	5.3.2 出版体系	176
3.12 INSPEC 数据库	143	5.3.3 著录格式	180
3.12.1 数据库简介	143	5.3.4 检索方法及实例	185
3.12.2 检索指南	144	5.4 中国专利文献及其检索	188
3.13 OCLC FirstSearch 基础组 数据库	144	5.4.1 中国专利文献	188
3.13.1 数据库简介	144	5.4.2 中国专利文献检索工具	189
3.13.2 检索指南	147	5.4.3 检索工具检索实例	190
3.14 中国资讯行	152	5.4.4 中国专利文摘数据库及检索	191
3.14.1 数据库简介	152	5.5 标准文献	195
3.14.2 检索指南	153	5.5.1 概况	195
3.15 全国报刊索引	156	5.5.2 中国标准文献	196
3.15.1 数据库简介	156	5.5.3 国际标准文献	197
3.15.2 检索指南	157	第6章 会议文献检索	199
第4章 学位论文文献检索	162	6.1 概况	199
4.1 国际学位论文检索工具	162	6.2 世界会议	200
4.2 PQDD 博硕士论文全文 数据库	164	6.3 ISTP 科技会议录索引	200
4.2.1 数据库简介	164	6.3.1 检索工具	200
4.2.2 检索指南	164	6.3.2 检索指南	201
4.3 南昌大学博硕士论文全文 数据库	167	第7章 电子图书检索	207
4.3.1 数据库简介	167	7.1 方正电子图书	207
4.3.2 检索指南	167	7.1.1 数据库简介	207
第5章 专利文献检索	168	7.1.2 检索指南	207
5.1 专利基本知识	168	7.2 超星电子图书	213
5.1.1 专利制度	168	7.2.1 数据库简介	213
5.1.2 知识产权和专利	168	7.2.2 检索指南	213
		7.3 NetLibrary 电子图书	226
		7.3.1 数据库简介	226

7.3.2 检索指南	226
第8章 学术文献检索与信息服务	232
8.1 信息服务	232
8.1.1 定题服务	232
8.1.2 代查代检	234
8.1.3 虚拟参考咨询	235
8.1.4 馆际互借与文献传递	237
8.1.5 科研产出能力及项目背景	

分析	240
8.2 文献检索与论文撰写	242
8.2.1 论文选题	242
8.2.2 查找文献	242
8.2.3 论文撰写	245
8.2.4 参考文献著录规范	246
8.2.5 投稿指南	248
参考文献	249

第1章 信息检索概论

1.1 基本概念

1.1.1 基础术语

(1) 信息 信息一词来自拉丁词源,意思是通知、报道或消息。在中国历史资料中,信息一词最早出自唐诗,是音信、消息的意思,一直沿用到20世纪中叶,其科学含义才被逐渐揭示出来。事实上,信息是事物的存在方式和运动状态的表征。信息对人类最重要的意义是它在人类认识和实践中的特殊作用。人类认识世界的过程,实际上就是不断地从外界取得信息和加工信息的过程;而人类改造世界的过程,则是把加工外部信息所得到的“主观”信息(表现为方针、政策、计划等等)反作用于外部世界的过程。因此,没有信息,就不能认识世界;没有信息,也不能有效地改造世界。

(2) 数字信息 通信系统传输和处理的对象,泛指消息和信号的具体内容和意义,通常须通过处理和分析来提取。数字信息就是以数字化的形式存储和检索的信息。

(3) 文献 文献是记录有知识和信息的一切载体,由四个要素组成:①所记录的知识和信息,即文献的内容。②记录知识和信息的符号,文献中的知识和信息是借助于文字、图表、声音、图像等记录下来并为人们所感知的。③用于记录知识和信息的物质载体,如竹简、纸张、胶卷、胶片等,它是文献的外在形式。④记录的方式或手段,如铸刻、书写、印刷、复制、录音、录像等,它们是知识、信息与载体的联系方式。

(4) 信息管理 是以信息组织为中心,注重静态对象;以文献信息为核心,注重加工、保存与外部形态整合,是对文献信息进行分类与检索、整序与整合、过滤与选择、应用与共享、分析与评价的过程。

(5) 知识与知识管理 知识是人类认识的成果或结晶。知识的特征有:隐性的——存在头脑中的经验、直觉和社会网络;镶嵌的——在程序、过程、系统和产品之中;显性的——在数据库、文件和表格中(专利权、程序指南、科学作品和图表)。知识管理是指将各种信息转化为知识,并将知识与人的需求和决策联系起来的过程。它以知识创新为中心,注重动态过程;以人和用户为核心,注重应用与共享和知识资源的增值。知识管理的目标是确保知识积累、知识获得、知识导航、知识开发、知识共享、知识创新和知识保护。

1.1.2 文献类型

1. 按文献载体形式分类

文献按载体形式分类,可分为印刷型、缩微型、电子型和声像型四种。

(1) 印刷型文献 以纸张为载体,以手写、印刷为记录手段而产生出来的一种传统的文献形式。其优点是便于阅读和流传,不受时间、地点和条件的限制;其缺点是存储密度太低,体积笨重,占储藏空间过多,不易管理和保存,难以实现自动输入和自动检索。

(2) 缩微型文献 以感光材料为存储介质,以缩微照相为记录手段而产生出来的一种

文献形式。它包括缩微胶卷、缩微胶片和缩微卡片,随着激光和全息照相技术的应用,又出现了超级缩微胶片和特级缩微胶片,最高缩小倍率已达 1/22500。一张全息胶片可存储 20 万页文献。其优点是体积小,存储密度高,传递方便,可以大大节省储藏空间;其缺点是不能直接阅读,需要借助缩微阅读机阅读。

(3) 电子型文献 是指经过数字化技术处理后存储在一定的特殊介质上并被计算机阅读和检索的文献资料。电子文献的基本特征是:文献的文本是以计算机可读数据的形式加以表示的;电子文献的信息载体是磁性或光学信息存储介质;电子文献经过格式化处理因而具有数据库结构;电子文献的使用要有相应的检索软件和其他编辑软件的支持并借助于计算机进行。其优点是存储密度高,存储速度快,可对所记录的信息进行各种处理,如转存、检索、传送、提取、检测和输出等;其缺点是使用费用较高。

(4) 声像型文献 又称为视听资料。这是一种以磁性材料和光学材料为存储介质,以电磁转换或光电转换技术为记录手段,直接记录声音、图像的一种文献形式。它包括:唱片、录音带、幻灯片、电影片、电视片、录像带、激光唱盘、激光电视录像盘、多媒体学习工具、程序设计工具等等。声像型文献的优点是存储密度高,内容直观,表现力强,易被理解接受,传播效果好。这类文献在帮助人们观察罕见的自然现象、探索物质结构等方面,具有独特的作用。

在上述几种载体中,印刷型文献是文献检索的基础检索工具,而电子型文献是文献检索的主体检索文献和发展方向。本书所涉及的范围以电子型文献和印刷型文献为主。

2. 按文献内容的加工与传递层次分类

按文献内容的加工与传递层次分类,可分为一次文献、二次文献和三次文献。

(1) 一次文献 是指作者以本人的研究成果为基本素材而创作(或撰写)的文献,一般一次文献记录的信息比较具体、详尽和系统化,大部分期刊论文、科技报告、专利说明书、会议论文等都是—次文献。习惯上—次文献也称为原始文献。

(2) 二次文献 是指文献工作者对一次文献进行加工整理之后所得到的产物,是为了便于管理和利用一次文献而编辑、出版和累积起来的工具性文献。它包括书目、索引、文摘等。二次文献的重要性在于它可以帮助人们查找一次文献内容,提供—次文献的线索。

(3) 三次文献 是指利用二次文献提供的线索,选用一次文献内容,而编纂出的成果。如百科全书、年鉴、手册、专著、教科书、述评、文献指南等等。

从—次文献到二、三次文献是一个从分散的原始文献经加工整理系统化的过程。一般说,一次文献是基础,是检索的对象;二次文献是检索一次文献的工具;三次文献是一次文献的浓缩。

3. 按文献出版形式分类

按文献出版形式分类,可分为图书、期刊(连续出版物)、会议文献、科技报告、标准文献、产品样本、专利文献、学位论文、档案文献、政府出版物等十大类型。也可将非书非刊的其他文献统称为特种文献,即图书、期刊和特种文献三大类。

(1) 图书 包括专著、教科书、参考工具书、丛书等。其特点是:内容比较系统、全面、成熟、可靠,有一定的新颖性,但编辑出版时间过长,传递信息的速度太慢,时间上难以满足获取最新科技信息的要求。

(2) 期刊 是一种周期性的连续出版物。它具有品种多、数量大、出版周期短、报道

(1) 磁性型文献 也称计算机阅读型文献,它是通过编码和程序设计把文献变成计算机可识别的数字语言,即机器语言,储存在磁带、磁盘等载体上。阅读时,再由计算机将存入的内容转换成人能直接阅读的自然语言,即通过数字技术将存储在物理介质上的信息转化为数字信息。它的特点是存储容量大,并能按一定的程序设计快速输出文献单元及知识单元。

(2) 光盘型文献 光盘是一种利用激光技术在特制圆盘上记录(写成)和再现(读出)信息的载体。具体来说,利用激光作为光源照射在光盘表面,使表面物质产生变化而记录信息;在读出信息时,也是利用激光在光盘上产生出不同强度的反射光而判断信息内容所代表的是“1”或“0”。光盘可分为记录模拟信息的视频式光盘和记录数字信息的数字式光盘两种,分别大致相当于磁记录中的调频式记录与数字式记录。视频式光盘适合于记录彩色或活动图像,而数字式光盘则适合于记录分辨率要求较高的图像。数字式光盘由于采用了自动纠错技术,带信息的光盘可以无损失地复制,其信息寿命接近于无限。

(3) 网络型文献 网络信息具有信息的时效性、内容的广泛性、访问的快速性、资源的动态性、搜索的相关性、组织的相对无序性。而网络型文献是指通过网络传播的并经过组织、制作成有序的数字信息文档。

1.2 信息检索与检索系统

1.2.1 信息检索

信息检索(Information Retrieval)是指将信息按一定的方式组织和存储起来,并根据信息用户的需要找出有关的信息的过程和技术。它的全称又叫“信息存储与检索”,这是广义的信息检索。狭义的信息检索则仅指该过程的后半部分,即从信息集合(库)中找出所需要的信息的过程,相当于人们通常所说的信息查询。在这里,存是指一种面向来自各种渠道的大量信息而进行的高度组织化的存储。而所谓取,就是面向随机出现的信息需求而进行的高度选择性的检索,要求能快速便利地查找与需求有关的信息。

信息检索的基本原理,可以用一句话把它最本质的部分概括为:对信息集合与需求集合的匹配与选择。即从用户需求出发,对一定的信息集合采用一定的技术手段,根据一定的线索与准则找出相关的信息。

1.2.2 检索系统

信息检索系统(information retrieval system):是利用一定设备和方法从某种载体(书本、胶卷、磁带或光盘、网络)上的文献、事实或数值记录集合中查找所需信息的系统。信息检索系统具有存储和检索两种基本功能。存储,即将搜集到的文献、数据进行分析、加工,给出检索标识和地址,组成卡片目录或记录在磁带或光学的载体上组成数据库。检索,即人工或自动地从书本式的检索工具或数据库中查找所需的信息。因此,信息的存储和检索是检索系统的两个不可分割的环节。

检索系统按其功能划分,可分为三种类型:目录检索系统、文献检索系统和事实检索系统。

1. 目录检索系统

目录检索系统是对出版物进行报道和对图书资料进行科学管理的工具,有出版发行目录、馆藏目录、资料来源目录等之分。根据文献出版类型,又可分为图书目录、报刊目录、

标准目录等。

馆藏目录反映图书资料的收藏情况,是用户经常使用的一种目录,以此来确定原(全)文的收藏地点。

馆藏目录可以是单个图书信息中心的,也可以是许多图书信息中心的馆藏的集合。后者称作联合目录,它可以反映一个地区、一个集团以至一个国家的图书资料收藏情况。目录系统的载体也是多样的,手工检索的目录有卡片目录和书本式目录。计算机检索的电子版目录,包括机读目录以及在网络上运行的联机公共检索目录(OPAC)。联机公共检索目录(OPAC)可提供联机目录查询,网络把许许多多图书信息中心连接起来,提供地区性的乃至全球性的目录检索服务。

2. 文献检索系统

传统的文献检索系统是采用对自然语言事先规范而形成的受控语言(如分类表、主题词表),来描述文献信息特征、生成概念及其概念标识系统,人们通过分类表中的分类符号或主题词表中的主题词(或叙词)作为控制检索的入口格式进行检索。

文献检索系统提供对参考文献的检索,检索结果往往是一些可提供研究课题使用的参考文献的线索,即来源信息。它是一种间接的相关性检索,即只能查出与用户提问相关的文献,供用户参考,而不能直接解答用户所提出的问题。

从文献检索到原文获取,往往需经过目录检索这一环节,从已检出的文献出处(如刊名及卷期号)等来源线索去查找馆藏单位,索取原文。

索取原文的方法有:

- 1) 由馆藏目录获得馆藏信息。
- 2) 联机公共检索目录(OPAC)可提供联机目录查询、目录检索服务,从而获得馆藏信息。

目录检索与文献检索系统总称书目检索,其数据库称为书目数据库,也称二次文献数据库,他们提供的信息是文献的线索,而不是文献全文。

3. 事实检索系统

事实检索系统包括对事实、数据和全文的检索。从广义上讲,事实和数据也是一种全文,只是内容比较特殊、比较简短的全文。事实检索提供事实、数据等信息的原始资料,是一种直接的确定性检索。

事实检索的手检工具有:词典、百科全书、年鉴、手册、指南、名录等。与之相对应的计算机检索系统称源数据库。它主要有以下四种类型:

(1) 数值数据库 专门提供数值方式表示的数据,直接提供解决问题所需的数据,是进行各种分析、统计、决策的重要工具。如统计数据库、财务数据库等。

(2) 文本-数值数据库 能同时提供文本信息和数值信息。如产品数据库、商情数据库等。

(3) 属性数据库 提供物理、化学、生物等科学数据。

(4) 全文数据库 提供文献的全文或其中的某些部分。如报刊全文库、法律法规全文库等。

1.2.3 检索系统的组织

检索系统是有序的文献信息集合。每篇文献都需要经过加工,把文献的特征描述下来,

成为一条著录款目,也称一条记录。将一个个记录按一定序列编排起来便组织成一个可供检索的系统。

文献检索系统按其检索手段划分有手检系统和机检系统,按其报道范围划分有多学科的综合性和专业性检索系统;按其文献来源划分有多种出版物类型检索系统和单一出版物类型检索系统。

1. 手工检索系统

根据著录方式不同,手工检索系统可分为四种:目录、题录、文摘、索引。

(1) 目录(catalog)是图书、期刊等单位出版物外表特征的揭示和报道,是按某种顺序编排的文献清单。目录通常是以一个完整的出版单位或收藏单位为著录的基本单位,即以文献的“本”、“种”或“件”(如一种期刊或一件专利)为报道单位。它对文献的描述比较简单,每条款目的著录项目有:著者/编者、书名/文献名和出版事项等。

(2) 题录是单篇文献外表特征的揭示和报道。题录通常是以一个内容上独立的文献单元(如一篇文章或书中的某一部分,甚至整本出版物)为著录的基本单位。每条款目的著录项目通常包括:篇名、著者和来源出处。

(3) 文摘是带有内容简介的、扩展了的题录。其著录项目比题录至少多一个文摘项。文摘也是以一个内容上独立的文献单元为著录的基本单位,它不仅描述文献外表特征,而且还包括文献的内容特征。文摘按揭示文献内容的详细程度可分为报道性文摘、指示性文摘和评论性文摘。

(4) 索引是对一组信息集合的有系统的指引,一般只起指引特定信息内容及其存储地址的作用。索引款目(条目)一般有两个著录项目:检索标识(标目)和存储地址。标识是索引款目所指示的文献某方面的特征,其属性值有著者姓名、主题词等。存储地址是指文献属性值所对应的特定信息内容在信息集合中的地址,多数是流水号(如文摘号)。不同的标识系统构成不同的索引,有主题索引、分类索引、著者索引、关键词索引、图书索引等等。比如主题索引就是以主题词作标识(目)。索引不能作为一个单独的检索系统存在,它是目录、题录、文摘检索系统中的一个重要组成部分,在检索系统中占有重要位置,它提供了从不同角度(检索点)对特定信息集合检索的途径。

2. 计算机检索系统

计算机检索系统由硬件、软件和数据库组成。就检索而言,检索系统是指数据库。其数字检索机制主要包括四个部分:①检索界面模块:接受用户检索要求,有一般检索界面和高级检索界面。②检索策略模块:将用户输入的检索要求编制成计算机可执行的规范化检索式。③检索执行模块:利用检索式检索索引数据库,并保证检索的速度和准确性。④检索结果组织模块:对检中记录的整理组织。

计算机检索系统有题录型、文摘型、指南数值型、全文型四种数据库。

(1) 题录型数据库 数据库存储某个领域文献的题录,包括文献的题名、作者、出处、主题词等。如:维普公司题录、文摘版数据库;德温特世界专利索引数据库等。

(2) 文摘型数据库 存储文献内容中主要部分的数据库,包括文献的题名、作者、出处、关键词、文摘等,基本上每一种书本式的检索工具都同时出版电子版,供用户在计算机系统上建立数据库。如:美国工程索引数据库、科学文摘数据库、化学文摘数据库等。

(3) 指南数值型数据库 数据库的记录主要是指关于一些机构、公司、企业、名人、

主要产品及其产量、价格、规格型号等信息的简单描述,通过这些数据库可以查到公司、机构的地址,产品目录、研究项目或名人简历等信息。如:万方数据公司的中国企业、公司及产品数据库(CECDB)、维普公司的中文期刊(引文索引版)数据库等。

(4) 全文型数据库 数据库存储文献内容的全文或其主要部分的数据库。如:中国学术期刊(全文版)数据库、Elsevier、Springer Link 等。

1.2.4 检索系统的构成

检索系统的构成一般是指手工检索系统和计算机检索系统的内容构成。

1. 手工检索系统

手工检索系统通常称之为书本式检索工具,一般由正文部分、辅助索引、使用说明和附录四个部分组成。

(1) 正文部分 作为手检工具的主体,正文部分是文献内、外部特征著录款目(条目)排成的有序集合。每一篇入存文献,只作一条款目,每条款目都有一个号码以与其他款目相区别。由于这一号码同时起排检顺序的作用,故称为顺序号或文摘号。不论什么检索工具,著录款目至少要表述文献的篇名、著者和文献来源(出处)等几个重要著录项目。大多数检索工具的正文部分都是按照学科分类体系的序列编排组织的,因此提供了分类检索途径(检索点)。

(2) 辅助索引 辅助索引是对正文单一线索检索方式的补充。索引的种类越多,检索的途径就越多。常见的有著者索引和主题索引。按出版时间划分,索引有:期索引、卷索引、年度和多年度累积索引。

(3) 使用说明 使用说明是用户使用某种检索工具的入门向导与指南,一般包括编制目的、学科范围、收录年限、著录方式和使用方法。

(4) 附录部分 其内容包括:引用期刊一览表、刊物名称的简称与全表、机构名称缩写表、分类表和主题词表等。分类表和主题词表是用户用于浏览、确定分类类目、类号和主题词的工具,分类号(类目名)、主题词作为检索点,用于正文或索引的准确检索。

2. 计算机检索系统

计算机检索系统由硬件、软件和数据库组成。数据库由文档、记录、字段组成。

(1) 字段(field) 书目数据库中基本的信息单位著录项目在数据库中称作字段。它是文献著录加工的最基本单元,描述的是文献外部特征和内容特征。字段包括:文摘字段、题名字段、著者字段等等。每一字段均有其标识符。

(2) 记录(record) 在数据库中的一篇文献称为一条记录,它是由若干个不同的著录单元(即字段)组成的。在数据库中每一条记录都有一个记录号,唯一地标识这条记录。

(3) 文档(file) 由某一类型的若干条记录组成的信息集合称之为文档。它是数据库的基本形式,包括顺排文档和倒排文档。

顺排文档指文档中的记录按序存放,记录的存取按物理顺序进行。书目数据库中的主文档通常是顺排文档,主文档与手工检索工具中的正文部分相对应,是数据库的主体。主文档像手工检索工具一样,需要辅助索引的配合即倒排文档的配合。

倒排文档是将每篇文献记录中的标识(如文献的主题词、题名、著者等)抽取出来,同时记录下该检索标识出现过的文献记录号,然后再按序组织起来成为可以用作索引的文档,这种倒排文档也称索引文档。如对著者字段,则将其所有记录字段中的著者名抽取出来,按序

组织成著者倒排档(即著者索引),以指引与特定著者有关的记录在主文档中的地址。倒排文档又可以分为主索引文档和辅助索引文档。

数据库是计算机信息检索系统的组成部分,它是在计算机的存储设备上合理存放的相互关联的数据的集合,通常,由存储信息记录及其索引的若干文档组成。不同的数据库,存储不同的主题内容、不同的时间范围和信息类型以及不同的标引方式,提供不同的检索范围和检索途径。因而掌握了数据库的特点可以自由灵活地运用检索技术,确保检索效果。

1.3 检索语言

检索语言(retrieval language)是用于描述检索系统中文献主题内容特征及外表特征和表达用户信息提问的一种专门语言。在文献检索的存储和检索过程中,检索语言起着重要的语言保障作用,是沟通文献存储和检索两个过程、标引人员和检索人员双方思想的桥梁。如果没有检索语言作为标引人员和检索人员的共同语言,就很难使标引人员对文献内容的表达(标引用语)和检索人员对同一内容文献需要的表达(检索用语)取得一致,文献检索也就无法进行。所以检索语言实质上是标引和检索之间的约定语言。使用检索工具必须掌握检索语言,如同使用计算机必须掌握计算机语言一样。掌握检索语言是掌握检索技能的基础。

检索语言作为文献存储检索系统的一个要素,在其中起着语言保证作用。基本功能是:

- 1) 对文献的情报内容(及某些外部特征)加以标引。
- 2) 对内容相同及相关的情报加以集中或揭示其相关性。
- 3) 对大量情报加以系统化或组织化。
- 4) 便于将标引用语和检索用语进行相符性比较。

1.3.1 检索语言的类型

检索语言就其组配方式来划分,可分为先组式检索语言和后组式检索语言;就其表达的文献特征来划分,可分为描述文献外表特征的检索语言

和描述文献内容特征的检索语言。描述文献外表特征的检索语言,最常用的有引文语言、题名(书名、刊名、文献篇名等)、著者、文献号码(入藏登记号、ISSN号、专利号、报告号等)。描述文献内容特征的检索语言,包括体系分类语言、标题词语言、叙词语言、关键词语言、单元词语言等。检索语言的类型如图1-1所示。

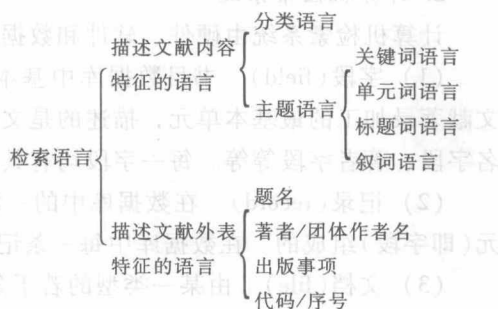


图 1-1

检索语言的种类虽然很多,但是在常用的手工检索工具中所使用的检索语言,不过是体系分类语言、人工语言、自然语言。不同的检索语言构成不同的标识和索引系统,以提供不同的检索点和检索途径。

1.3.2 体系分类语言

1. 体系分类法构成原理和基本作用

分类是区别事物及其相互关系的一种思想方法,是人们思维活动中的一种本能。体系分

类法所依据的原理,就是对知识概念的划分。一类事物除了具有共同的属性以外,还有许多不同的属性,即同类事物并不完全相同,不同的属性可以使同类事物进一步划分,所以“类”是可以划分的。

体系分类是一种直接体现知识分类的等级概念的标识系统。它根据一定的观点,以科学分类为基础,文献内容的学科性质为对象,运用概念划分与概括的方法,按照知识门类的逻辑秩序,从一般到具体,从简单到复杂,层层进行划分。每次划分,就产生许多类目;逐级划分就产生许多不同级别的类目。所有不同级别的类目,层层隶属,形成了一个严格有序、线性的知识门类等级制体系。每个类目都用分类号作标识,每个分类号都是表达特定知识概念的词汇,即体系分类语言的语词。

分类号的作用:

1) 表示类目在分类体系中的位置。

2) 表示类目的排列顺序。

2. 体系分类法的特点

(1) 体系分类法的优点:

1) 以学科的衍生、发展为基础进行分类,符合人们认识事物的习惯,容易被人们熟悉、掌握、使用。

2) 较好地体现了学科的系统性,反映了事物的隶属、平行、派生关系,有利于人们从专业、学科的角度检索文献。

3) 采用国际上广泛应用的阿拉伯数字、拉丁字母组成检索标识,通用性强,简单易记,人们容易理解和接受。

4) 具有直线性序列特点,从而为文献的分类组成与编排存储发挥了独特的功能。

(2) 体系分类法的缺点:

1) 采用从上位类到下位类逐级展开细分而形成的纵向线性关系,难于反映当代科学技术交叉渗透而出现的多元知识空间,不适于标引和检索主题因素复杂的多维概念问题。

2) 各级类目、类号规模宏大,全都一一列举,不能及时全面地反映日新月异的科学技术动态。因此,体系分类法的类目设置总是落后于科学技术的发展,难于检索出边缘学科、新兴学科的内容。

3) 人类知识无限专深而又广博,而分类表的类目设置总是有一个限度,这一对矛盾就造成了不能准确检索专指度较高的课题。

3. 《中国图书馆图书分类法》(简称:《中图法》)

一部完整的分类法通常由“分类表”、“辅助表”、“使用手册”三大部分构成。

分类表是类分文献和组织文献的依据;辅助表是为了帮助用户从主题的角度迅速、准确地查得分类号。《中图法》配置有相应的辅助工具《中国图书馆图书分类索引》,是提供按主题途径检索《中图法》分类表的辅助工具,该索引按汉语拼音字母顺序排列。使用手册主要包括:体系结构的编制原则与方法;类分文献的基本规则;大类编制结构与分类方法;同类书排列等方面。

分类表是类分文献和组织文献的依据。它一般分成:编制说明、基本部类、大纲、简表、详表等几部分。

(1) 编制说明 内容包括分类表的编制经过,依据的编制原则,部类及大类的设置和

排列次序的理由,对各种分类问题的处理方法、使用方法等等。其功能是帮助用户对分类表形成一个系统而全面的认识。

(2) 基本部类 又叫基本序列,是对文献最基本的划分依据。

《中图法》采用五分法,将人类全部知识划分为:马克思主义、列宁主义、毛泽东思想;哲学;社会科学;自然科学;综合性图书。

(3) 大纲 也叫基本大类,是在基本序列的基础上进一步展开成基础大纲,构成分类表的一级类目。它的作用是让用户对整个分类体系形成一个最基本的整体概念。《中图法》遵循从一般到具体、从简单到复杂、从低级到高级,以及学科之间的内在联系,排列成22大类,见表1-1。每个大类下还有二级类目,见表1-2。

表 1-1 基本大类

A 马克思主义、列宁主义、毛泽东思想	N 自然科学总论
B 哲学	O 数理科学和化学
C 社会科学总论	P 天文学、地球科学
D 政治	Q 生物科学
E 军事	R 医药、卫生
F 经济	S 农业科学
G 文化、科学、教育、体育	T 工业技术
H 语言、文字	U 交通运输
I 文学	V 航空、航天
J 艺术	X 环境科学
K 历史、地理	Z 综合性图书

表 1-2 “T 工业技术” 大类二级类目

TB 工业技术	TL 原子能技术
TD 矿业工程	TM 电工技术
TE 石油、天然气工业	TN 无线电电子学、电信技术
TF 冶金工业	TP 自动化技术、计算技术
TG 金属学、金属工艺	TQ 化学工业
TH 机械、仪表工业	TS 轻工业、手工业
TJ 武器工业	TU 建筑科学
TK 动力工程	TV 水利工程

(4) 简表 即基本类目表,展现分类的骨架。大型分类表的正文类目很多,作线性排列时,极不易全部掌握。因此简表即成为一种过渡表,通过它就能够比较迅速地掌握整个分类表的总体内容及列类的分布状态。

简表的功能在于:其一,在查表时由简表入手,再转查详表,掌握分类体系的脉络,加快掌握并熟练地使用整个分类表的过程;其二,简表作为简略分类表使用。

(5) 详表 又称主表,即分类表的正文。详表由类号、类目、注释组成。类号使用不同的字号和字体表示相互隶属或并列关系;类目采用不同字号、字体和齐行、缩行表示类目之间的隶属、等级、并列的关系。详表是分类标引检索文献的依据。

《中图法》的详表排列是以基本大类分别展开,在每个大类和二级类目之前,列出“基本类目”,然后再排列详细的类号、类目、注释。这样,有助于用户迅速而准确地了解该类的内容,确定出所需的类号。

1.3.3 人工语言和自然语言

人工语言也称规范化语言,是指检索词来自文献或用户提问并受到文献检索系统控制(规范和管理)的一类检索语言的总称。人工语言是表达一系列概括文献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,用以对文献内容和情报需要进行主题标引、逻辑分类或特征描述。

所谓自然语言,是指作者的书面用语,可以减少概念间转换产生的误差,检索入口词多,操作简单、方便、灵活,也适合专业人员之外的广大用户群,可以大大缩短检索工具的时差,从而保证文献报道与检索的及时性。但是,如果选词不加严格控制会降低查准率。同时,由于自然语言对多义词也基本不加控制,往往使相关主题内容的文献分散,从而造成漏检。受控语言与自然语言存在的这种互逆相关性,恰好说明它们在网络环境中兼容、整合的必要性。在文献检索过程中,检索途径、检索语言、检索标识以及检索工具的辅助索引之间的关系是密不可分的,如图1-2所示。

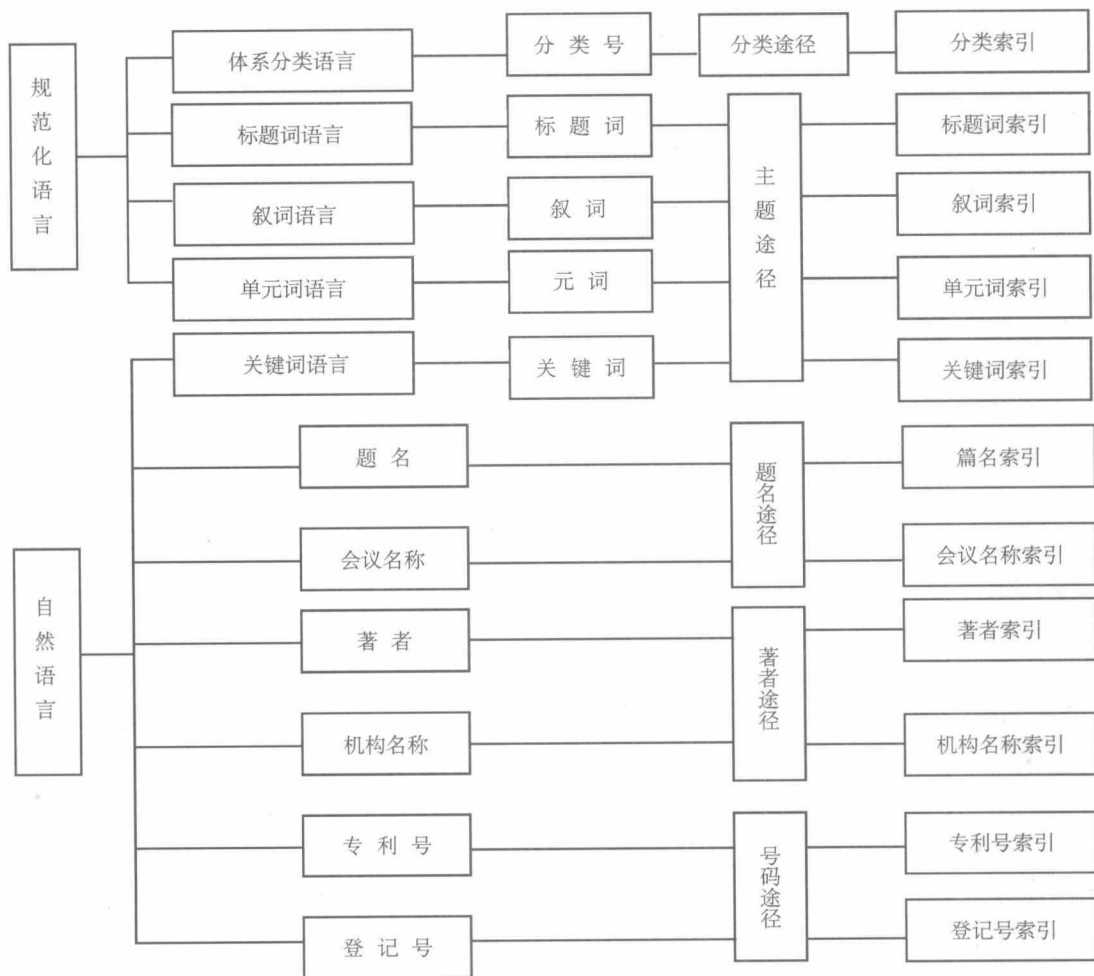


图 1-2