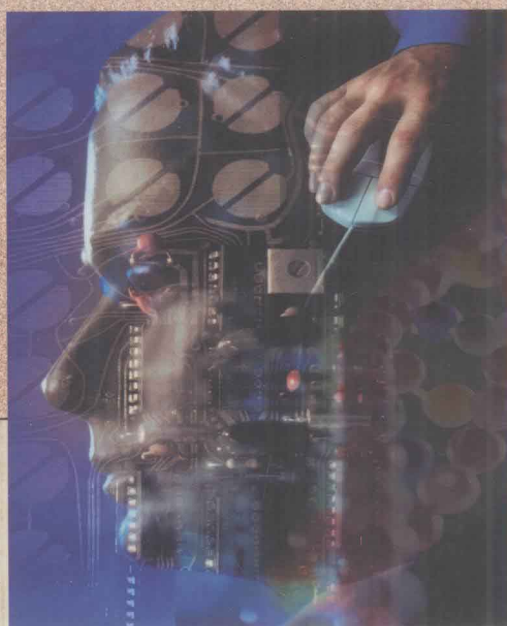


YIXUE WENXIAN JIANSUO

(修订版)

医学文献检索

余 鸣/主编



安徽大学出版社

图书在版编目(CIP)数据

医学文献检索(修订版) / 余鸣等编著. —合肥:安徽
大学出版社, 2002. 8

ISBN 7-81052-590-5

I. 医… II. 余… III. 医学—情报检索—医学院
校—教材 IV. G252.7

中国版本图书馆 CIP 数据核字(2002)第 061481 号

医学文献检索(修订版)

余 鸣 主 编

出版发行	安徽大学出版社	印 刷	合肥远东印务有限责任公司
	(合肥市肥西路 3 号 邮编 230039)	开 本	787×1092 1/16
联系电话	总编室 0551-5107719	印 张	14 75
	发行部 0551-5107784	字 数	340 千
电子信箱	ahdxchps@mail. hf. ah. cn	版 次	2002 年 8 月第 1 版
责任编辑	徐 建	印 次	2005 年 1 月第 2 次印刷
封面设计	孟献辉	经 销	新华书店

| ISBN 7-81052-590-5 / G·153 | | 定 价 24 00 元 | |

如有影响阅读的印装质量问题,请与出版社发行部联系调换

前 言

在信息时代,谁掌握了获取信息的原理、方法和技术,谁就拿到了通向成功的钥匙。在高校,信息检索课早已成为一门重要的课程,深受广大学生们的欢迎。在医学院校,医学文献(信息)检索课从 20 世纪 80 年代中期始亦普遍开设,并编写出版了一些有影响的教材。我们曾在 1999 年编写出版了《医学文献检索》(安徽大学出版社出版)教材,经过几年来的教学检验,我们发现书中存在一些问题,尤其是书中内容结构与快速发展的现代信息检索技术不相适应的问题,使我们深感有对该书进行修改的必要。经过多位老师的合作,终于使《医学文献检索》(修订版)得已问世。

本书的作者均为长期从事医学信息检索的教学和研究人员,他们既具有一定的信息检索的教学经验,又具有较强的计算机检索、网络信息检索的实际操作能力。

本书共分 6 章,重点介绍了国内外医学信息检索系统,并用大量篇幅突出介绍了计算机光盘数据库、镜像数据库检索和 Internet 医学信息检索。书中附有大量的图片和检索实例。

该书编写章节分工:

余鸣:第一章,第三章第四节,第五章第三节,第六章第一、二、五、六节;

朱康玲:第三章第二、三节,第四章第一、二节,第五章第二节,第六章第三节;

熊慧平:第三章第一节,第五章第一节,第四章第三节;

李桂芳:第二章第一节(三),第三章第五节,第四章第五节,第五章第四节,第六章第四节;

杨军:第二章第一节(一)、(二),第二节,第四章第四节。

该书的出版得到了安徽医科大学教材科的大力支持,在此表示感谢。

由于作者水平有限,书中难免有错漏之处,恳请广大读者指正。

编 者

2002 年 6 月 15 日

目 次

第一章 概论.....	(1)
第一节 文献的基础知识.....	(1)
一、信息	(1)
二、知识	(1)
三、情报	(2)
四、文献	(2)
第二节 文献检索的基本原理.....	(4)
一、文献检索	(4)
二、文献检索系统	(4)
三、文献检索原理	(4)
四、检索语言	(5)
第三节 检索工具.....	(7)
一、手检工具	(8)
二、机检工具的特点	(9)
三、机检工具类型	(9)
第四节 文献检索的基本方法和途径	(10)
一、文献检索的基本方法.....	(10)
二、文献检索途径.....	(11)
第二章 中文医学文献检索工具(印刷版)	(14)
第一节 检索国内医学文献的检索工具	(14)
一、《中文科技资料目录》(医药卫生).....	(14)
二、《中国医学文摘》.....	(19)
三、《中国科学引文索引》.....	(19)
第二节 检索国外医学文献的检索工具	(23)
一、《国外科技资料目录》(医药卫生).....	(23)
二、《国外医学》.....	(24)
第三章 国外主要医学文献检索工具(印刷版)	(26)
第一节 美国《医学索引》	(26)
一、概况.....	(26)
二、《IM》收录范围	(26)
三、编排体系和著录格式.....	(27)
四、医学主题词表.....	(28)
五、检索途径及示例	(40)
六、刊名、人名的缩写和音译方法	(42)

第二节 荷兰《医学文摘》	(48)
一、概况	(48)
二、编排结构及著录格式	(50)
三、检索途径与方法	(52)
四、检索策略	(53)
五、引用期刊表	(53)
第三节 美国《生物学文摘》	(53)
一、概况	(53)
二、编排结构与著录格式	(53)
三、检索途径及方法	(57)
第四节 美国《化学文摘》	(59)
一、概况	(59)
二、CA 的著录项目	(59)
三、CA 的组织编排	(60)
四、检索途径与方法	(63)
第五节 美国《科学引文索引》	(69)
一、概况	(69)
二、编排结构	(69)
三、索引及著录格式	(70)
四、检索途径和方法	(74)
五、检索注意事项	(76)
第四章 国内主要医学文献数据库	(77)
第一节 中国生物医学文献数据库(CBMdisc)(光盘版)	(77)
一、概况	(77)
二、数据库结构	(77)
三、数据库检索途径与检索方法	(78)
第二节 中文生物医学期刊数据库(CMCC)(光盘版)	(89)
一、数据库简介	(89)
二、数据库检索途径和检索步骤	(90)
第三节 中国学术期刊(光盘版)	(94)
一、概述	(94)
二、中国期刊网专题全文数据库检索方法	(95)
第四节 万方数据库(镜像)	(99)
一、数字化期刊子系统	(100)
二、科技信息子系统	(103)
第五节 《中国科学引文索引》数据库(光盘版)	(108)
一、概况	(108)
二、检索途径及方法	(109)
三、显示与输出	(109)

四、检索举例	(109)
第五章 国外主要医学文献数据库	(116)
第一节 MEDLINE(光盘版)	(116)
一、概况	(116)
二、MEDLINE CD-ROM 数据库结构	(117)
三、SilverPlatter MEDLINE CD-ROM 数据库操作系统:WINSPIRS	(120)
四、MEDLINE 数据库检索方法举例	(134)
第二节 《Biosis Previews》(镜像)	(136)
一、概况	(136)
二、检索主界面	(137)
三、数据库检索途径及方法	(139)
第三节 美国《化学文摘》(光盘版)	(151)
一、概况	(151)
二、《CA》光盘数据库结构	(151)
三、《CA》基本检索方法	(153)
四、检索结果的输出	(158)
五、CA 光盘版检索常见问题及技巧	(159)
第四节 美国《科学引文索引》(光盘版)	(160)
一、概况	(160)
二、检索功能	(161)
三、显示	(162)
四、打印和存储	(162)
五、检索举例	(163)
第六章 Internet 与医学信息检索	(169)
第一节 Internet 概述	(169)
一、什么是 Internet	(169)
二、Internet 的发展	(169)
三、与 Internet 连接	(170)
四、Internet 的主要功能	(170)
第二节 环球网(WWW)	(171)
一、WWW 简述	(171)
二、Web 浏览器	(171)
三、Web 信息查询原理	(172)
四、Web 检索工具类型	(172)
五、Web 检索工具	(173)
第三节 OCLC	(175)
一、概况	(175)
二、数据库简介	(175)
三、OCLC FirstSearch 检索	(176)

第四节 PubMed 数据库(MEDLINE 数据库的网站)	(181)
一、概况	(181)
二、收录范围	(181)
三、检索原理	(182)
四、PubMed 检索功能	(182)
五、检索举例	(188)
第五节 分子生物数据库 Genbank 检索	(192)
一、生物信息学概述	(192)
二、生物信息数据库的种类	(193)
三、分子生物数据库 NCBI 检索	(197)
第六节 生物医学专业网站介绍.....	(214)
一、医学专业网站	(214)
二、公共卫生专业网站	(224)

第一章 概 论

医学信息检索与利用是医学信息学的一个分支课程,是一门实践性强、应用面广的学科。学习本课程的目的,在于提高学生的信息素质,掌握医学信息检索的基本原理和基础知识,学会使用检索工具,并能熟练、准确、快捷地从检索系统中获取医学信息;通过学习,提高学生自我更新知识和独立从事科学研究的能力,把握国内外医学研究的最新动态,善于利用已有的科学研究成果,能在一个更新的高起点上进行研究和创造。

第一节 文献的基础知识

一、信息

信息可以认为是生物体或有一定功能的机器通过感觉器官或相应的设备同外界进行交换的一切内容的总括。信息在自然界、人类社会以及人类思维活动中普遍存在,是物质的一种基本属性,是物质存在方式及其运动规律的外在表现。不同的运动状态和特征会产生不同的信息反映,包括社会信息、生物信息等。在医学上,各种体征的出现和变化是反映疾病的信息,各种化验结果的数据也是反映疾病的信息。信息是无形的,但它是客观存在的,是认识事物发展的基础。人类发展的历史是在不断获取信息、认识信息、传递和利用信息、创造信息的过程中,通过信息认识世界与改造世界的历史。在医学上,人们通过各种疾病的不同信息来认识千差万别的疾病。

信息、物质和能源一起构成人类的财富和资源,成为现代科技的三大支柱。物质提供的是材料,能源提供的是动力,信息奉献给人类的则是知识和智慧。

信息是人类社会不可缺少的,其作用愈来愈重要。通常所说的“信息社会”的特征是:①在信息社会里起决定作用的不是资本而是信息和知识,信息已成为生产力、竞争力和科技成就的象征;②生产价值的增长主要靠知识和信息;③从事信息产业的人员占相当大的比例。

二、知识

知识是人们通过信息对自然界、人类社会以及思维活动规律的认识与掌握,是人的大脑通过思维重新组合的系统化信息的集合。医学知识是对人体生命、健康、疾病现象本质规律的认识,它来源于实践,通过长期实践、积累、优化、系统化而逐渐形成,因此它是医学信息的一部分。只有系统化的信息才是知识。

三、情报

情报是一种活化的信息和知识,是一种动态的信息和知识,它能被利用、被活化,否则它仍然是知识、信息的客观存在。信息要成为情报,一般要经过选择、综合、分析和研究加工的过程,即经过知识化的过程才能成为情报,因此它具有三种基本属性:知识、传递和效益。知识是情报的实体,传递是情报的表现形式,效益是情报的结果。医学情报就是人类在同疾病作斗争中对医药信息经过综合、筛选、逻辑思维、重新组合的系统化知识。

信息、知识和情报三者的关系:信息包含知识,知识是信息的一部分,情报包含信息和知识的特征,是活化的知识,能为人们所利用。其逻辑关系是:信息>知识>情报,而信息、知识、情报是文献的实质性内容。

四、文献

(一)文献的定义

凡属于人类的知识用文字、图形、符号、声频、视频等手段记录保存下来,并用以交流传播的一切物质形态的载体,都统称为文献。国家标准局于1983年7月2日颁布并于1984年4月1日实施的中华人民共和国国家标准《文献著录总则》中明确规定了“文献”一词的定义:“记录有知识的一切载体。”用一定方式记录在一定载体上的医药卫生方面知识称为医学文献。

(二)文献的种类

文献按照其外在的表现形式和内涵等,可以分成很多类。

1. 按文献载体形式分为

(1)纸张型。自纸张和印刷术发明以来,纸张型文献直到今天仍是主要的文献形式。其特点是符合人类传统的阅读习惯,实用方便。

(2)缩微型。是指采用照相技术,将文献按一定的比例,缩小存储在感光胶卷或平片上,通过专门的阅读机进行阅读。其特点是存储容量较纸张型大,保存期比纸张型长,但需要一定的设备进行阅读。

(3)声像型。包括录音、幻灯、电影信息。其特点是可以闻其声,观其行,给人以直接感受。

(4)机读型。主要指的是计算机阅读型。随着计算机及网络技术的发展,计算机阅读型的文献越来越受到人们的青睐。它以数字形式将文献存储在磁、光等介质(如磁盘、磁带、磁卡、光盘)上,通过本地计算机或远程通信传输的网络计算机系统来进行阅读。其中有文本(text)型、多媒体(multimedia)型、超文本(hypertext)型。

文本型文献主要是以数字形式记载文字信息,让用户通过计算机阅读。文本型的文献也可进一步分成题录型、文摘型和全文型。题录型的文献,通常只提供文献的著者、题名、原文的出处(原文登载的期刊名,发表的年、卷、期页等)。随着计算机的处理能力和存储容量的不断提高和扩大,机读型进一步发展到了文摘型的数据库文献,即除了提供用户题录外,还可检索文献的概要,即文摘。20世纪70年代以前所指的文献检索,大部分指的是对这些

类型的文本获取。当网络技术、超大容量计算机等飞速发展时,全文型文献计算机检索便应运而生。

多媒体型的文献是以数字形式包括计算机视频、音频等技术记载图形、影像、动画,并让用户既可通过计算机阅读,又可与计算机进行互相问答的信息利用方式。

而超文本(包括超媒体)则是由于计算机网络、特别是 WWW 技术的发展而产生的一种通过文本或图像的关键词或图标链接文件的形式,让用户不必考虑信息的来源或分类,自由地在网络中寻找相关信息的一种全新的计算机信息阅读的方式。

计算机机读型文献的特点是存储容量很大,传递迅速,便于检索,且效率高。

2. 按文献发布的类型分

(1)图书。指一些记录的知识比较系统、成熟的文献,是掌握一门学科的基本资料。一般都有固定的装帧,并出版发行。如教科书、丛书、专著、全集、会议论文集等,以及词典、百科全书、指南、手册等参考工具书等。

(2)期刊。指一些记录的知识比较新颖、所含信息密度比较大的连续出版物,一般都有固定的名称。如有关自然科学的综合性期刊《中国科学》、《自然杂志》等;医学学科的专业性期刊有《中华医学杂志》、《中国药理学通报》等;学报有《安徽医科大学学报》等;检索性期刊有美国《化学文摘》(Chemical Abstracts, CA)、美国《生物学文摘》(Biological Abstracts, BA)、美国《医学索引》(Index Medicus)等。期刊通常刊登了能够反映学科领域最新的理论、方法、技术等论文(journal article)、综述(review)、病例报告(case report)等各种文献信息。

期刊是科研课题工作的主要文献源、信息源、情报源。医学科研工作者依靠期刊来及时跟踪最新的国际国内的研究动向、把握科研的主动权。

(3)资料。指无法归入图书或期刊的文献,如科技报告、学位论文、专利说明书、标准文献、会议文献、政府报告等。这些文献一般不公开出版,普通图书馆也不收藏。但是这些文献有许多反映了最新的科学研究和技术发展情况以及国家的法规、标准定义等信息,也是医学科研的重要信息源。

3. 按揭示信息内容可分为

(1)一次文献。一次文献就是原始文献,是作者根据自己的工作和研究成果而写成的文章,也称原始论文。一次文献包括:期刊论文、研究报告、会议录、专利说明书、学位论文等。一次文献的内容有创新性,记录着科学技术的新发现、新发明、新理论、新见解,是重要的情报源。目前它有以下特点:①数量激增。有资料报道,近 20 年的科学文献的量相当于人类历史几千年来总和,并以每 10 年翻一番的速度增长。②文种繁多。根据世界卫生组织(WHO)的最新统计,医学文献目前在全世界发现有几个文种。③发表分散。由于学科的分支越来越细,一篇关于免疫学的文献可从肿瘤学、分子生物学、医学工程等多种期刊、杂志上获得。④老化加快。科技文献的寿命一般为 5 年到 10 年,而医学文献的老化速度更快,其“半衰期”一般为 5 年。

(2)二次文献。是指对原始信息进行加工、组织和整理而形成的便于管理和查找利用原始信息的工具,也称“检索工具”。二次文献是查找一次文献的线索,是检索文献时不可缺少的工具。它主要反映文献的来源、出处和内容概要,包括文献中的题名、著者、主题,注明原文的出处,以及收藏文献的图书馆或机构。二次文献包括:目录、索引、文摘等。

(3)三次文献。三次文献是利用二次文献并在其指导下,对大量一次文献,进行综合分

析研究加以浓缩和提炼而形成的文献,是对知识的第三次加工。三次文献一般是由专家撰写,水平较高,专业性强。

三次文献包括综述、评论、述评、进展、动态、年鉴、专著、指南等。

第二节 文献检索的基本原理

一、文献检索

要正确理解文献检索这一概念,首先需了解什么叫情报检索。情报检索是指将情报按照一定的方式组织和存储起来,并按照情报用户的需要查找出有关情报的过程。情报检索分为文献检索、数据检索和事实检索。文献检索是将文献按一定方式组织存储在某种载体上,并利用相应的方法或手段,从中找出符合需要的文献的过程。文献检索是最基本的检索方式,数据检索与事实检索,一般也要通过文献检索来实现,所以国内外不少人把文献检索与情报检索等看待。三种检索类型只是检索要求不同,其检索原理并无本质上的区别:如某用户需要“有关艾滋病防治”方面的文献,这是属于文献检索范畴的问题;数据检索则属一种确定性检索,即直接提供所需的确切的数据;事实检索也属于一种确定性检索,即直接提供用户所需的确切的非数值性数据(即事实)。

二、文献检索系统

文献检索系统是存储经过加工而有序的情报信息,并能提供文献检索的一种工作系统,因此,它具有存储和检索两种基本职能。文献检索必须通过检索系统来实现。

文献检索系统分为手工检索系统和机械检索系统两大类。手工检索系统包括卡片式和书本式的目录、索引和文摘;机械检索系统包括光电检索系统和电子计算机检索系统等。文献检索系统具体地说,就是通常所说的手工检索工具和机械检索工具。它是由检索文档、技术设备、作用于系统的人和逻辑语义工具(检索语言、文献标引规则等)所组成的一个完整有序的系统。

三、文献检索原理

文献检索原理是建立在检索系统基础上的,因为检索系统为文献检索提供了检索途径,两者之间有密切的相互联系。文献检索系统具有存储和检索两种职能,存储是文献的输入过程,检索是文献的输出过程;存储是检索的基础,检索是存储的目的。通俗地讲,文献检索原理的实质就是,按存储文献的相同方法从检索系统中获得文献的过程。

(一)标引

是利用一定的标识符号,按照一定的方法和规则,对文献中具有检索意义的特征(包括外表特征,如书名、著者、出版年等;内容特征,如分类、主题等)进行分析、著录的全部处理过程。标引,顾名思义是标记和指引,目的是起到指引的作用。

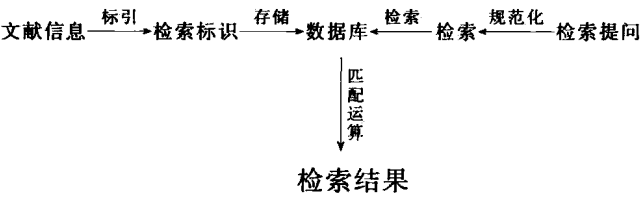
使用分类法,从分类的角度标引,从中选择恰当的分类号来表达文献的中心问题或主

题,称分类标引;使用主题词表,从主题的角度标引,从中选择适当的主题词来表达文献的中心问题或主题,叫主题标引。每篇文献在输入检索系统或检索文献之前,都要进行文献标引。

(二)检索标识

为了确切地表达文献内容特征和外表特征,常使用一种便于记忆与书写的约定符号,这些符号分为号码标识、语词标识和代码标识三大类。号码标识(分类号)是利用数码、字母等按照一定的编排方法组合后作为标识,其系统性好;语词标识(检索词)利用自然语言的语词经过规范化处理后作标识,直观性强。标识要符合惟一性、通用性等要求,为此要对标识进行规范化处理。简言之,检索标识是在进行文献标引时,对文献内容进行分析,从中找出能代表其内容实质的词汇或符号,作为检索和存储的根据。例如“中药丹参对心肌细胞钾通道的作用机理探讨”一文,其检索标识可选用“丹参”、“心肌细胞”和“钾通道”。

文献存储是对文献内容进行主题分析形成主题概念,再利用检索语言(如分类语言或主题语言)进行标引,把主题概念转换成检索标识(分类号或主题词)后,按一定方法排序、输入到检索系统之中的工作。目前大部分系统的组成如下:



文献检索原理是揭示如何将文献的有关特征输入到检索系统中,又如何从检索系统中输出的过程。只要掌握了检索原理,就能得心应手地获得所需文献。

四、检索语言

(一)检索语言

检索语言是根据文献检索的需要而创制的人工检索语言,又称文献存储与检索语言、标引语言、索引语言等。

文献检索语言的实质就是表达一系列概括文献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或专业文献中精选出来并予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(如代表化合物的多种代码)等等。它们可用于对文献的内容进行主题标引、逻辑分类或特定信息的揭示和描述。例如,《汉语主题词表》、《医学主题词表》、《中国图书馆分类法》等,各自都可称为文献检索语言的一个语种。

文献检索语言既是汇集、组织、存储文献的手段和工具,也是使标引人员和检索人员取得共同理解、实现交流的语言。文献信息检索语言的控制性文本(如特定的图书资料分类法、主题词表、关键词法的禁用词表等)既有供检索系统标引人员使用的工作用文本,也有供

检索人员(读者)使用的简略文本。

(二)检索语言的种类

文献检索语言可分为描述文献外部特征的语言和描述文献内容特征的语言,它们是文献信息检索语言的主体与核心。

1. 外部特征的检索语言

描述文献外部特征的检索语言是依据文献的外表特征,如用文献题名、著者、文献序号等作为文献存储的标识和用文献检索提问的出发点而设计的索引语言。其应用主要有:

(1)题名索引系统。以文献上记载的书名、刊名、篇名等作为标识的字顺索引系统,如书名目录(索引)、刊名目录(索引);作为检索工具组成部分的“引用期刊一览表”、篇名索引等。

(2)著者索引系统。以文献上署名的著者、译者、编者等的姓名或机关团体名称作为标识的字顺索引系统。如著者目录(索引)、专利权人索引等。

(3)序号索引系统。以文献特有的序号作为标识的索引系统。如专利号索引、科技报告号索引、技术标准号索引。

(4)引文索引系统。引文索引系统是根据科学文献末尾所附参考文献及所引用文献这一外表特征,揭示科学论文之间引用与被引用关系而编制的索引系统。如美国的《科学引文索引》(Science Citation Index)。

2. 内容特征的检索语言

描述文献内容特征的检索语言按其构成原理,可分为分类检索语言、主题检索语言和代码检索语言三大类型。

(1)分类检索语言。在分类语言中,应用比较普遍的是系统分类法,又称为体系分类法或等级体系分类法。体系分类法是一种直接体现知识分类的等级制概念标识系统。其主要特点是按学科、专业集中文献,从知识分类的角度揭示多种不同文献在内容上的区别和联系,提供以学科分类为出发点的文献信息检索途径。

等级体系分类语言的“语词”就是它的类目。一部等级体系分类法由成千上万个类目构成。一个类目表示具有某种或某些共同属性的一“类”事物。类是可分的,同类事物可根据事物间的不同属性,或者说,根据该类事物所表达的概念的外延再划分为子类。被划分的类称为上位类,划分出来的类称为下位类。上位类与下位类之间是隶属关系,各子类之间是并列关系,互称同位类。全部类目遵守逻辑分类规则,层层划分,构成具有隶属、并列关系的概念等级体系。

体系分类法的类目用一种具有自然顺序的标记符号来代表,这就是分类号。分类号的主要作用在于明确各个类目之间的先后顺序,以便排列目录和组织藏书。同时,分类号也是分类检索语言的语词。它代表一个类目,也就表示这个类目的涵义,分类号系统也在一定程度上表达各类之间隶属、并列关系。

国内外有多种著名的图书分类法都是等级体系分类法,如:美国《国会图书馆图书分类法》(Library of Congress Classification,简称 LC)、《杜威十进分类法》(Dewey Decimal Classification and Relative Index,简称 DC 或 DDC)、《国际十进分类法》(Universal Decimal Classification,简称 UDC),以及我国的《中国图书馆图书分类法》(简称《中图法》)。

(2)主题检索语言。主题检索语言中应用较多的是叙词语言和关键词语言。

叙词又称为主题词。叙词法的主要特点:①叙词是经过严格规范化的词和词组,保证语词与概念的一一对应;②叙词可用于进行概念组配检索;③采用参照系统显示非叙词与叙词之间的等同关系以及某些叙词之间的相互关系;④采用体系分类法的基本原理编制叙词分类索引(范畴索引)和等级索引(词族索引),采用与关键词法类似的方法编制叙词轮排索引,以从多方面显示叙词间的相互关系。

一部叙词表通常还是动态性的,随着科学的发展及文献中用词的变化,叙词表一般会不断增删修订,定期更新。

我国的《汉语主题词表》和美国的《医学主题词表》(Medical Subject Headings)都是叙词表。

关键词是为适应计算机自动编制索引的需要而产生的。所谓关键词(Keyword)是指出现在文献的标题(篇名、章节名)或文摘、全文中,能表达文献实质内容的,或者能被人们作为检索入口的关键性专业名词术语。例如:一篇文献的篇名是“癌基因和抑癌基因在肺癌研究中的应用和发展趋势”,其中“癌基因、抑癌基因、肺癌”三个词能表达该文题的主要含义,可作为关键词,并可作为检索入口的重要词语,而“研究、应用、发展趋势”只起到辅助描述作用,不能作为检索入口词,因而不是关键词。关键词通常取自文献原文,不作规范。

关键词法的最大优点是语词直接采用出现在最新文献中的专业自然语言,一些在科学发展中最新出现的专门名词术语能及时地进入索引系统。同时,计算机编制关键词索引还具有速度快、时差短的优点。

关键词因用词不作规范,使自然语言中大量存在的同义、多义、近义、同形异义等词因用词不规范,使同一主题的文献常因用词而分散,造成漏检的可能性较大;同时,若平均每篇文献标引的关键词较多,虽然可以减少漏检,但误检的可能性会增大。此外,关键词法不显示词间等级关系和相关关系,增加了查全文献的难度。

用关键词语言编制的索引有带上下文的关键词索引和不带上下文的关键词索引两大类,前者如美国《生物学文摘》的(1997年以前的)“Subject Index”,后者如美国《化学文摘》的“Keyword Index”。

(3)代码检索语言。代码检索语言是用代表事物的代码作为标识系统的索引语言。具体的应用有化合物分子式索引、环状化合物的环系索引系统等。

第三节 检索工具

检索工具是报道、存储和查找文献线索的工具,即对文献信息的内容特征和外部特征进行分析筛选,使用一定的检索语言,按照一定的学科或主题范围规范成一定的文献信息条目,采用适当的编排方法,组织在一起而形成的二次文献。其中,文献信息条目是指描述文献外部特征(题名、著者、出处等)和内容特征(主题词、分类号)的记录单元。被标引了的文献特征叫检索标目,它是存储和检索文献的入口词。

一般说来,检索工具由文献信息库和逻辑语义工具两大部分组成。文献信息库是检索工具的主体,在计算机光盘检索中也称为数据库。完整的文献库由说明、目录、正文、索引和附录五部分组成。说明部分是检索工具编制者对用户所提供的使用指南,其内容一般包括

该检索工具的收录内容、适用范围、著录格式等,一般放在检索工具的最前面。目录部分多出现在用分类检索语言编制的检索工具中,其实它相当于一部分分类简表。检索工具的正文部分是浓缩后的文献信息条目集合体,由原文的题录或文摘所组成。此外为了标明文献信息条目的地址和排检顺序,每条文献信息条目都对应着一个文摘号(顺序号)。正文部分反映检索工具的主体内容,用户根据它就可判定原文内容是否符合提问需要并找到获取原文的线索。索引部分是以正文为基础,从中选出有检索意义的外部特征或内容特征并用某种检索语言标引后所形成的检索标识集合。

一部好的检索工具除了能从正文部分提供一条检索途径外,为了多角度、全方位地揭示检索工具的功能,还必须编制各种各样的辅助索引,索引的种类越多,提供的检索途径就越多,用户使用起来就越方便。附录部分是检索工具的必要补充,大型检索工具一般都设有附录,其内容包括引用期刊一览表、刊名缩写全称对照表、机构名称字首缩略语表等。

逻辑语义工具即检索语言、规则等,它是沟通文献标引人员与检索人员双方思想的媒介,其实就是检索人员和标引人员的约定语言,只不过它独立于具体的检索工具之外,为若干检索工具所通用。

一、手检工具

所谓手检工具就是无须借助任何辅助设施即可使用的检索工具。这类工具都是传统的印刷型检索工具,因此用户使用起来灵活、方便、费用低,便于阅读;但这类工具存储密度低、占空间大、管理困难。

手工检索工具的类型很多,按其不同的特征,可划分出不同的类型。若按出版形式划分,可分为卡片式和书本式检索工具,书本式又可分为期刊、单卷和附录式三种形式;若按收录范围划分,可分为综合性检索工具和专业性检索工具;若按著录方式划分,可划分为题录、文摘和索引三种形式。

(一)题录

题录是描述文献信息的外部特征、并按一定的方式编排起来的文献条目的集合体。文献的外部特征包括题名、著者、出处、文种、年卷期、页次等。其特点是报道及时,数量大,但揭示文献的深度不够、准确性差。

题录著录的对象可以是整体文献(比如一本书),也可以是单篇文献(比如一篇文章)。对于著录整体文献的题录称为“目录”或“书目”。目录按照编制目的和社会职能,还可划分为国家书目、出版社与书商目录、馆藏目录、联合目录和专题文献目录等。

题录性检索工具很多,如《全国新书目》等。

(二)文摘

文摘是对文献内容作实质性描述的文献条目,有时也称摘要、提要或简介。文摘只对文献的内容做简洁、准确的摘录,是原始文献浓缩后的产物,它既包含文献的外部特征,又包含文献的内容特征,且编写中不得加入编写人员任何主观性见解和解释性说明,它必须基于原文、忠于原文。

由于文摘不仅著录了文献的外部特征,还描述了文献的内容特征,所以文摘性检索工具

首先在揭示报道文献的深度及实用性等方面都优于题录性检索工具。其次,文摘在揭示和报道最新文献信息的同时,在一定程度上还能起到替代原文的作用,从而大大加速了文献检索的进程,起到事半功倍之效。此外,文摘还能为用户消除语言障碍,因为在一种检索工具中,无论原文语种如何,其文摘都被翻译成同一种语言。

文摘的种类很多,如报道性文摘、指示性文摘、评论性文摘等。

(三)索引

索引是将文献信息中某些具有检索意义的文献特征标识(如人名、地名、序号、主题词、分类号、分子式等)按一定的顺序加以排列,并注明其所在文献信息中的位置(如文摘号、页码、版次等),以便检索相应文献信息的检索工具。它的特点是用户利用简单、方便,但也和题录性检索工具一样未能反映文献内容。

索引在这三种检索工具中是应用最广的一种,它不仅有单独出版的索引刊物,还有许多文摘题录性检索刊物之后附有各种辅助索引,如主题索引、著者索引、专利索引等。辅助索引的应用很普遍,现已成为检索工具不可分割的一部分。

二、机检工具的特点

机检工具是指检索者必须借助各种设备才能使用的检索工具。这类工具大多以缩微型和机读型的形式出现。这类检索工具与手检相比具有鲜明的特点:

- 检索效率高。机检速度快,机检查找一个课题,一般需要的时间较短,检索效率高,节省人力。相比之下,手检望尘莫及。

- 检索范围广。手检一般只能在本馆、本地区查找,所受限制太多。而机检利用终端设备就地即可检索远方的大型计算机系统的信息,从而打破了时空界限,实现了跨越国家和地区的世界范围内的检索。

- 检索内容全。只要是数据库中有的,机检可不受文献内容交叉、出版分散等限制而迅速准确地检出;另一方面机检可减少中间环节,一次性检索到原文。而手检一般局限在本馆、本地区所收藏的文献上,容易造成漏检。

- 检索费用高、灵活性差。这是因为机检必须借助专门的设备。相比之下,手检灵活、方便、费用低。

三、机检工具类型

机检工具若按容纳信息载体来划分,可分为磁带、缩微制品和光盘等几种;若按数据库类型划分,又可划分出以下几种形式:

(一)书目数据库

又称文献数据库,实际上是文摘、索引期刊和图书目录实现计算机化的产物,故每个数据库一般都有相应的书本式检索工具或卡片式目录。该数据库一般能指引用户到另一信息源中索取原文或其他某些细节。所以“文献出处”至关重要。

书目数据库记录由一些表示文献特征的字段和子字段组成,这些字段往往作为用户检索文献的入口词。字段分主题性和非主题性两种:主题性字段有标题、叙词、分类代码、文

摘、标引词及其他;非主题性字段有著者、单位、期刊代码、出版年月、语种、更新日期等。但对大多数用户来说,这些数据记录还不能满足他们的需要,必须借助“出处”再查一次文献。

目前,书目数据库在所有数据库中占主体,几乎覆盖了全部专业领域的各类文献。如大多数光盘数据库和联机数据库均属此种。

(二)目录数据库

也称指南数据库或事实数据库,主要是对一些公司、团体、企事业单位、名人、化学物质等作一些指示性描述。通过该数据库可检索到公司、团体、企事业单位的地址和电话传真、产品种类、研究项目、物质名称等记录内容。目录数据库与书目数据库的主要区别是:目录数据库不收文献资料,只收指南性资料;目录数据库常对公司名、人名、专用名等其他非主题性数据字段进行检索。

(三)数值性数据库

这是一种专门以数值表格方式提供各种数据服务的源数据库。这些数据均是来自原始文献的统计数据、调查数据或经过处理的其他数据集合。该数据库以数值数据为主,有时也包含文字,但文字是用来定义数字所必需的最小量的文字,这时就称文本-数值数据库。

用户查找数值数据库一般可一步到位,不必再查阅其他信息源。

(四)全文数据库

它是数据库记录中含有文献全文或其中主要部分的一种源数据库。用户利用该库可直接检索原文,或检索全文中的段、节、章等内容。如新闻消息、法律法规、期刊论文、辞典或百科全书等都可以制成全文数据库。

(五)多媒体数据库

前面所提到的各种数据库,如书目、目录、数值和全文数据库都是文本型,它们均没有将声音、图形、动画等各种各样的媒体信息融为一体;多媒体数据库则突破这一局限,集文本、声音、图像等信息于一体,使信息交流不局限于通过文字形式,还可以通过上述各种媒介进行“全方位”的,即“多媒体”的交流。该数据库与文本数据库的一个最大不同在于它能够将多种信息组装在一起,形成有声有色、图文并茂的系统。

第四节 文献检索的基本方法和途径

一、文献检索的基本方法

(一)常用法

又称工具法,即利用各种检索工具查找文献的方法。常用法有三种: