

YIXUE XINXI JIANSUO  
JIAOCHENG



# 医学 信息检索 教程

◎董建成 周晓政 主编

东南大学出版社

# 医学信息检索教程

主 编 董建成 周晓政

副主编 张政辉 朱佳鸣

编 者 (按姓氏笔画排序)

王云峰 申飞驹 朱佳鸣 沈 放

吴季平 张红萍 张志美 张政辉

周晓政 陈馥瑛 梅 谊 董建成

蒋 葵 程 杰 蔡忆宁

东南大学出版社

南 京

## 内 容 提 要

本书详细介绍了医学信息检索的基本原理、方法和发展趋势。内容包括:信息检索基础;经典的医学检索工具书和数据库,如《中目医》和 CBMDisc、美国《医学索引》和 MEDLINE 及 PUBMED 等;网络基础知识、数字图书馆、电子期刊、目前热门搜索引擎和电子论坛的使用,医疗卫生工作者通过网络能够进行会诊并获得国内外专家的指点和帮助;及时反映最新的检索状况,如 CNKI 最新网络服务平台 KNS3.5 版检索系统、中国医院知识仓库(CHKD)的介绍和科技查新的最新规范,资料截至于 2002 年 7 月。为满足医务工作者适应经验医学向循证医学转变所需的知识信息,详细介绍了循证医学的概念、国内外发展情况、循证医学的文献类型以及如何及时获取有关疾病诊断、病因、治疗、预后的最新权威文献信息,指导临床实践。

本书数据库检索介绍详细,将枯燥的信息检索语言融于具体数据库的检索实践中,通过具体的检索实例使读者能够掌握检索的精华,便于读者自学,适合于大学生、研究生、医务工作者以及与医药学信息相关的人员作为教科书和参考书。

### 图书在版编目(CIP)数据

医学信息检索教程/董建成等主编. —南京:东南大学出版社,2002.8

ISBN 7-81050-971-3

I. 医... II. 董... III. 医药学—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(2002)第 054878 号

东南大学出版社出版发行  
(南京四牌楼 2 号 邮编 210096)

出版人:宋增民

江苏省新华书店经销 南京京新印刷厂印刷

开本:700mm×1000mm 1/16 印张:21 字数:435 千字

2002 年 8 月第 1 版 2002 年 8 月第 1 次印刷

印数:1~10000 定价:26.00 元

(凡因印装质量问题,可直接向发行科调换。电话:025-3792327)

# 序

信息检索能力的培养为人们的终身学习打下基础。

信息检索、分析、组织和利用的能力,即当前倍受国内外关注的“信息素养”(Information Literacy)。美国大学与研究型图书级协会指导委员会于2000年1月18日批准的《高等教育信息素养能力标准》认为,一个具备信息素养的人应达到如下目标:

“确定所需信息的范围;有效地获取所需的信息;鉴别信息及其来源;将检索出的信息融入自己的知识基础;有效地利用信息去完成一个具体任务;了解利用信息的过程中所涉及的经济、法律和社会问题,合理、合法地检索、搜集和利用信息。”

由董建成教授主编的《医学信息检索教程》,正是针对医学院校的本科生、研究生和临床医生继续教育的对医学信息检索的需求,较为全面地分析、研究、介绍了医学信息检索的基本原理、方法和发展趋势。从章节体例上来看,本书的“信息检索基础”等8个章节可谓是精简明了。其内容紧紧围绕着医学类的信息检索与利用这一主体,书中介绍了网络信息检索、中外医学信息检索、特种文献检索、医学参考工具书使用、科技查新、医学信息检索与循证医学等。这对于在校大学生、研究生的学习和参考具有适用性和针对性;对于医学界的研究人员和临床医务人员来说,亦可作为手中必备的工具书或参考书。本书书后的8个附录,如,MeSH 范畴表主要类目,专题副主题词表及树状结构表,EM 分册系列表,BA 主要概念字顺表,CA 类目表,科技部、教育部、卫生部认定的查新机构名录等对于读者也有很好的参考和使用价值。

本书可以说是新时期作者从事医学信息科研与教学工作的新的成果与经验总结。

董建成同志1982年毕业于南通医学院临床医学专业,获医学学士学位。毕业后从事医学图书情报工作,同时承担医学科研方法教学工作。1984年,国家教育部发文要求高等学校开设文献检索课程后,专门从事医学信息检索教学与研究的工作。1988年到北京大学信息管理系攻读科技情报专业研究生,其专业研究方向是医学信息检索和医学情报计量学。在北大获理学硕士学位。多年来,他在医学专题信息检索和网络信息方面的研究与开发利用,以及在文献计量学领域的研究成绩斐然。目前,除了专业教学外,还承担了国家卫生部“医学信息检索”网络课件的研制。近20年来发表专业论文30多篇。主编和参编的专著有5种。

面向在校大学生讲授信息检索课,已成为我国大学图书馆用户教育的一大特色。自检索课程在部分高校开设以来,教育主管部门、学校的教务处和图书馆做了

很多工作,信息检索课程教材的出版也如雨后春笋,发展很快。董建成教授主编的《医学信息检索教程》,以一个具有医学专业人员的专业背景眼光来总结、介绍信息检索的技能和方法。这正是本书的可贵之处,全书结构严谨,重点突出,层次清楚,通俗易懂,实用性强,医务工作者和高校医学类大学生、研究生若能细细品味,定能起到事半功倍的效果。

张厚生

2002年8月18日于东大

# 前 言

《医学信息检索教程》为我国高等医学院校本科生的重要课程之一,是医务工作者适应医学信息化十分必要的一门课程。随着社会经济的高速发展,科学技术的不断进步,新的信息技术层出不穷,人们的信息意识、信息观念也在不断更新。在信息检索领域,新的检索内容、检索技术不断产生,传统的信息媒体和检索工具、检索方式在不断发生变化。为了适应这些新的变化和发展,适应我国高等医学教育的改革要求,我们组织了有关医学信息检索课程的现任教师和研究人员,在参考和借鉴国内外信息检索技术的基础上,结合多年来的教学实践,编写出这本《医学信息检索教程》。

《医学信息检索教程》共分8章,内容包括三部分:第一部分(第1~2章),简要介绍了医学信息检索的基础理论、基本概念,医学信息类型,信息检索的方法、途径和步骤以及常见的现代信息检索技术和网络检索方法。第二部分(第3~6章),以中文生物医学文献数据库(CBMDisc)、中国知识基础设施工程(CNKI)和中文科技期刊数据库(VIP)为例,介绍了中文医学信息检索的方法和途径;详细介绍了MEDLINE、EMBASE、BA、CA、IPA、SCI、CC等国外主要生物医学数据库的最新检索技术;同时,对与生物医学有关的会议文献、科技报告、标准文献、专利文献、学位论文等特种文献信息的检索和参考工具书中事实与数据信息的检索方法也作了必要的介绍。第三部分(第7~8章),主要介绍了循证医学和医学科技查新的有关知识,为医学生和医务工作者进一步综合利用医学信息检索技术提供了参考。

在本书的编写过程中,得到程纯、李文华等同志的关心和支持;本书汲取了国内外许多专家学者的有关研究成果,同时,承蒙张厚生教授为本书作序,在此一并致谢。

限于水平,书中难免有欠妥之处,殷请广大师生和读者不吝赐教,惠予指正。

编 者

2002年6月于东南大学榴园

# 目 录

<b>第一章 信息检索基础</b> .....	(1)
第一节 信息与文献.....	(1)
第二节 信息检索.....	(4)
第三节 信息检索技术 .....	(16)
<b>第二章 网络信息检索</b> .....	(22)
第一节 因特网基础 .....	(22)
第二节 搜索引擎 .....	(34)
第三节 数字图书馆 .....	(44)
第四节 电子期刊 .....	(53)
第五节 电子论坛 .....	(63)
第六节 网络医学信息检索策略 .....	(70)
<b>第三章 中文医学信息检索</b> .....	(74)
第一节 主要印刷型检索工具 .....	(74)
第二节 中国生物医学文献数据库(CBMDisc).....	(80)
第三节 中国知识基础设施工程(CNKI) .....	(93)
第四节 中文科技期刊数据库(VIP) .....	(105)
<b>第四章 国外医学信息检索</b> .....	(110)
第一节 美国《医学索引》与 MEDLINE .....	(110)
第二节 荷兰《医学文摘》与 EMBASE .....	(133)
第三节 美国《生物学文摘》(BA)与 BIOSIS .....	(140)
第四节 美国《化学文摘》(CA)与 CA on CD .....	(152)
第五节 美国《国际药学文摘》(IPA) .....	(186)
第六节 美国《科学引文索引》(SCI) .....	(190)
第七节 美国《近期期刊目次》(CC) .....	(214)
<b>第五章 特种文献检索</b> .....	(224)
第一节 会议文献.....	(224)
第二节 科技报告.....	(228)
第三节 标准文献.....	(231)
第四节 学位论文.....	(233)
第五节 专利信息.....	(238)

<b>第六章 医学参考工具书</b> .....	(248)
第一节 参考工具书概述 .....	(248)
第二节 医学参考工具书的利用 .....	(252)
第三节 电子版参考工具书 .....	(258)
<b>第七章 医学信息检索与循证医学</b> .....	(265)
第一节 循证医学概述 .....	(265)
第二节 循证医学实践 .....	(267)
第三节 循证医学信息检索 .....	(268)
<b>第八章 科技查新工作</b> .....	(277)
第一节 科技查新概述 .....	(277)
第二节 科技查新程序 .....	(280)
第三节 科技查新的质量评判体系 .....	(287)
第四节 控制科技查新质量的其他措施 .....	(292)
第五节 科技查新争议 .....	(292)
<b>附录</b> .....	(294)
附录一 拉丁字母 - 日文字母音译对照表 .....	(294)
附录二 俄文字母 - 英文字母音译对照表 .....	(294)
附录三 MeSH 范畴表主要类目 .....	(295)
附录四 专题副主题词表及树状结构表 .....	(299)
附录五 EM 分册系列表 .....	(313)
附录六 BA 主要概念标题等级表 .....	(315)
附录七 CA 类目表 .....	(318)
附录八 科技部、教育部、卫生部认定的科技查新机构名录 .....	(321)
附录九 本书重要字词索引 .....	(323)
<b>参考文献</b> .....	(325)



# 第一章 信息检索基础

信息检索(Information Retrieval)是指信息的有序化识别和查找的过程,即人们根据特定的信息需求,采取科学的方法,应用专门的工具,从浩瀚的信息海洋中迅速、准确地获取所需信息的过程。早期的信息检索,人们主要根据文献的特征,用手工方式实现。以计算机为核心的信息技术,开辟了信息处理与信息检索的新纪元,计算机从处理数字信息发展到处理字符信息、静态和动态的图像信息,乃至声音信息等,不仅拓展了信息检索的领域,丰富了信息检索的内容,而且极大地提高了信息检索的速度,信息检索已成为当今科学研究、经济活动和社会生活的一个组成部分并发挥越来越大的作用。

## 第一节 信息与文献

信息是许多学科广泛使用的概念,在不同的学科领域,对其有着不同角度的解释。但人们普遍认为信息与能源、材料科学并列,构成现代社会的三大支柱。

在信息检索领域,一般将信息理解为关于现实世界事物存在方式或运动状态的反映。例如,某患者,年龄 58 岁,男,身高 1.72 m,体重 69 kg,体温 37.2℃,患有糖尿病,这些都是关于患者的信息,是患者存在状态的反映。

信息有许多重要的特征:信息来源于物质和能量;信息是可以感知的;信息是可以存储的;信息是可以加工、传递和再生的。这些特征构成了信息的最重要的自然属性。作为信息的社会属性,信息已经成为社会上各行各业不可缺少的重要资源之一。人类获取、积累并利用信息是认识和改造客观世界的必要过程。借助信息,人类才能获得知识,才能有效地组织各种社会活动。因此,信息是人类维持正常活动不可缺少的资源。

几乎和信息同样广泛使用的另一个概念是“文献”。所谓文献,是指记录有信息的一切载体,即采用文字、图像、符号、声频、视频、编码等方法记录人类信息的各种载体,它们具有传递信息的功能。

由此可见,信息与文献之间存在着固有的联系:文献是信息的载体,信息则是文献的内涵。文献表示了信息,而信息通过文献形式表示出来才容易被人们理解和接受。

### 一、信息的类型

#### 1. 文字信息

文字是人们为了实现信息交流、通信联系所创造的一种约定的形象符号。广义的文字还包括各种编码,如 ASCII 码、汉字双字节代码、国际电报与单元代码以及计算机中的二进制数字编码等。

## 2. 图像信息

图像是一种视觉信息,它比文字信息直接,易于理解。人工创造的图像,如一张纸、一幅画、一部电影、大自然的客观景象等都是抽象或间接的图像信息。随着多媒体技术的发展,各类图像信息库将会极大地丰富人类生活。

## 3. 数值数据信息

数值数据是“信息的数字形式”或“数字化的信息形式”。狭义的数据是指有一定数值特性的信息,如统计数据、气象数据、测量数据以及计算机中区别于程序的计算数据。广义的数据是指在计算机网络中存储、处理、传输的二进制数字编码。文字信息、图像信息、语音信息以及从自然界直接采集的各种自然信息均可转换为二进制数码,网络中的数据通信、数据处理和数据库等就是广义的数值数据信息。

## 4. 语音信息

人讲话实际上是大脑的某种编码形式的信息转换成语音信息的输出,是一种最普遍的信息表现形式。音乐也是一种信息形式,是一种特殊的声音信息,它是通过演奏方式表达丰富多彩的信息内容的。

# 二、文献的类型

## 1. 印刷型

印刷型文献又称纸质型文献,是指以手写或印刷技术为主要手段、以纸张为信息记录载体的文献。其优点是可以直接阅读,携带方便,是目前人类信息交流活动中最常用的工具。与现代信息载体相比,印刷型文献存储信息密度低,占用收藏空间大,不易长期保存,难以实现自动化输入和自动检索。

印刷型文献可分为三类:

(1) 图书 图书(Book)通常提供比较系统、成熟的知识,一般包括专著、教科书、论文集和参考工具书等。专著是对某一个专题有较深入的研究和独到见解的学术著作,如《心血管药理学》、《休克》等。教科书是某个专业或学科的研究总结,反映较成熟的专业理论,具有严格的系统性与逻辑性,内容可靠性强,是医学生和医学工作者进行专业学习的主要医学文献。论文集是由多位作者的论文或会议论文、报告等汇编而成的出版物。参考工具书是供日常工作、学习或写作中随时查阅用的一类图书,其内容有序,便于查考,主要包括字典、词典、年鉴、手册、名录、图谱、百科全书等(详见第六章第一节)。

(2) 期刊 期刊(Journal)也称“杂志”,是指具有相对固定的刊名、编辑机构及版式装帧的连续出版物,如美国的“Science”(科学)、英国的“Nature”(自然)、我国的《中华医学杂志》等。期刊的内容通常是能够反映学科领域最新的理论、方法、技术

的论文(Journal Article)、综述(Review)、病例报告(Case Report)等。期刊论文包括研究报告、论著等,是反映最新科研成果,具有学术性、创新性和科学性特点的医学信息。综述是综合描述某一专题或学科在一定时间内的研究现状和进展的文献,其综合性强,权威性高,能够直接反映专业领域内科研的动向和进展。

(3) 特种文献 又称非书非刊资料,包括除图书、期刊以外的其他出版物,常为不定期出版,多数具有连续性。其特点是:数量大、种类多、内容广、参考价值大。

- 政府出版物:是指国家各级政府部门及其所属机构出版的文献信息资料,主要包括社会科学与自然科学两大类。其中行政文件,如讨论会记录、各种法令、外交文件、统计数据占大多数,科技资料数量相对较少。

- 会议文献:是指在国内外学术团体举行的专业会议上发表的论文或学术报告,其特点是信息传播速度快,反映研究成果新。会议文献主要通过会议论文摘要、论文集、期刊特辑或增刊等形式予以刊载。

- 专利文献:专利(Patent)是指受到法律保护的技术发明。专利文献是指发明人向政府部门(专利局或称知识产权局)递交的、说明自己创造的技术文件,同时也是实现发明所有权的法律性文件,包括专利说明书、专利公报等,具有新颖性、创造性、实用性。

- 科技报告:是指各学术团体、科研机构、高等院校的研究报告及其研究过程的记录,其理论性较强,它反映了某一专业领域科研进展和动态。但其保密性强,通常难以获取。

- 技术标准和规范:又称标准文献,是有关产品或工程质量、规格、生产过程、检验方法的技术文件,具有一定的法律约束力。主要包括技术标准、技术规范、操作规程、准则、术语等。标准分为国际标准、区域标准、国家标准 and 行业标准四大类。

- 学位论文:毕业生为取得博士、硕士、学士学位而撰写的学术性研究论文。
- 其他:如报纸、手稿、内部刊物、病历档案、技术资料、产品样本,等等。

## 2. 缩微型

缩微型信息是以感光材料记录信息的一种载体,如缩微胶卷、缩微胶片、计算机存取载体的输出胶片(Computer Output Microfilm, COM)等。缩微型信息具有载体体积小、存储信息密度高、成本低、便于保存等优点,但使用时,必须借助缩微阅读设备进行阅读复印。

## 3. 声像型

声像型信息又称为视听型信息,是一种记录声音、图像信息的载体,如照片、录音带、录像带、幻灯片、影视片、视听光盘等。声像型信息的载体形式可以让人们通过视觉、听觉感受到直观、形象、生动、逼真、丰富多彩的信息世界。

## 4. 电子型

电子型文献也称机读型文献、数字型文献,是采用电子手段并以数字形式存

储、利用计算机及现代通讯方式提供信息的一种新型信息载体,如光盘数据库、网络数据库、电子图书、电子杂志、电子地图等。数字型信息载体的问世是信息时代的重要标志,它改变了传统书刊的物理形态,开辟了一种新的信息载体的信息传播渠道,极大地提高了信息的传递速度,加速了社会信息化的进程。与传统信息载体相比,其优点是信息容量大,传递速度快,便于检索且效率高。目前,电子型文献与印刷型文献已共同成为当前科学信息的两大主流载体。常见的电子型文献有:

(1) 数据库 数据库(Database, DB)可以直观地理解为存放数据的仓库,只不过这个仓库是在计算机的大容量存储器上,如磁盘数据库、光盘数据库(CD-ROM)、联机数据库、网络数据库等。数据库中的数据可以是数字,也可以是文字、图形、图像、声音等,虽然数据库有多种表现形式,但它们都是经过数字化处理后存入计算机的。

(2) 网络文献 网络文献的出版、传递、检索和利用是通过 Internet 得以实现的,通常利用 WWW(信息浏览)、FTP(文件传输)、Telnet(远程登录)、Gopher(信息查找)、Archie(文件名查询)、USENET(网络新闻)、E-mail(电子邮件)等方式检索 Internet 上的各种各样的信息。

(3) 印刷型文献的数字化 主要是将印刷型文献数字化后,制成可供计算机阅读、检索和利用的电子出版物,主要有电子图书、电子杂志、电子地图,等等。

另外,有学者按对信息加工深度的不同将文献划分为一次文献、二次文献和三次文献。一次文献即原始文献,是作者以生产或科研成果为依据而创作的文献,如专著、期刊论文、研究报告、学位论文、发明专利等。二次文献是根据一次文献的内容和外表特征进行加工整序后的文献,如目录、索引、文摘、书目数据库、搜索引擎等,常被视为信息检索工具的主体。三次文献是对一次和二次文献进行综合、分析后编辑而成的文献,如综述、评论、科技动态、进展、指南等。

## 第二节 信息检索

广义的信息检索包括信息的存储和信息的检索,往往又称为“信息存储与检索”(Information Storage and Retrieval)。信息的存储主要是在一定专业范围内的信息选择基础上进行信息特征描述、加工并使其有序化,或建立数据库,以便在检索时借助一定的设备与工具,从中查找出所需的信息。存储是检索的基础,检索是存储的反过程。在现代信息技术的条件下,信息检索从本质上讲,是指人们从任何信息系统中高效、准确地查找到自己所需的有用信息,而不管它以何种形式出现,或借助于什么样的媒体,此即狭义的信息检索。本书所讲的信息检索主要指的是后者。

### 一、信息检索系统

信息检索系统是根据社会发展需要和为达到特定的信息交流目的而建立的一

种有序化的信息资源集合体。它通常是一个拥有选择、整理、加工、存储、检索信息的设备与方法,并能够向用户提供信息服务的多功能开放系统。由下列要素构成:

#### 1. 信息资源

信息资源是系统存储与检索的对象。它可以是全文信息,也可以是题录、索引或文摘;它可以是文字信息,也可以是图形、图像、数值、数据或语音信息。

#### 2. 设备

即实现信息存储与检索活动的一切设备。如手工检索的卡片、印刷型检索工具、计算机、交换机、服务器、通讯网络、软件,等等。

#### 3. 方法与策略

包括检索语言、标引方法、信息的组织与管理方法、信息的检索策略与技巧等。

#### 4. 人

人是检索系统的能动因素。充当信息与用户媒介的检索人员,将随着社会网络化程度的不断提高而逐步退出系统,由具有自主检索能力的最终用户取而代之。

## 二、信息检索类型

### 1. 按信息检索的对象分类

(1) 文献检索 文献检索就是从大量的文献集合中查找出符合特定需要的相关文献的过程。一般是先查找出相关文献的线索,如题录、文摘等,然后进一步查寻原始文献进行阅读参考。文献检索的结果是有关某课题或特定需要的一组相关性文献。

(2) 数据检索 数据检索是以特定的数值型数据为检索对象的检索过程。包括各种统计数字、图表、化学结构式、计算公式等,如胰岛素的理化常数、结构式、常用剂量等。

(3) 事实检索 事实检索是利用特定的参考工具书或事实型数据库查找出能够直接解答某一提问的事实。例如,什么是基因工程?何谓生物芯片?何人何时在何处首先提出了人类基因组计划?

综上所述,数据检索、事实检索是一种确定性检索,其检索结果可以直接回答有或无,正确或错误;而文献检索是一种相关性检索,其检索结果只提供与之相关的文献以供参考,不直接回答用户提出的问题。文献检索是信息检索的一个重要组成部分,科技人员进行信息检索的过程中,通常以文献检索为主。

### 2. 按信息组织的方式分类

(1) 目录检索 目录检索是指通过卡片式目录、书本式目录、机读目录(Machine Readable Catalog, MARC)或联机公共检索目录(Online Public Access Catalog)查询单位出版物(如一本书、一种杂志、一件专利)的名称、著者、出版事项等文献外表特征的过程,供人们了解出版或收藏机构是否拥有所需的图书、期刊等出版物的情况。

(2) 题录检索 题录检索类似于目录检索,但其检索结果不是单位出版物,而

是单位出版物中单篇文献的外表特征。如美国的《医学索引》(Index Medicus, IM)、《中国生物医学文献数据库》(Chinese BioMedical Disc, CBMDisc)等。

(3) 文摘检索 文摘检索是在题录检索的基础上,增加了反映文献的主题范围、目的、方法、结果等内容特征的摘要。有利于引导用户阅读原文,节约阅读时间,确定所获文献与用户需求的相关程度,如 Medline 数据库、美国《生物学文摘》(BA)、荷兰《医学文摘》(EM)等。

(4) 全文检索 全文检索是用户根据特定的需要,从存储有整篇文章乃至整本图书的信息检索系统中获取全文或有关章节信息的过程。利用全文信息检索系统,还可以进行各种信息的频率统计和内容分析。随着计算机容量与运行速度的不断增大与提高,全文检索正迅速由最初的法律、文学领域扩展到几乎所有的学科和专业领域。

(5) 超文本检索 超文本的基本组成元素是节点(Nodes)和节点间的逻辑链接链(Links),每个节点中所存储的信息以及信息链被联系在一起,构成相互交叉的信息网络。与传统文本的线性顺序不同,超文本检索强调中心节点之间的语义联接结构,依靠系统提供的复杂工具作图示穿行和节点展示,提供浏览式查询。其检索模式是从“哪里”到“什么”。而传统的文本检索系统则强调文本节点的相对自主性,其检索模式是从“什么”到“哪里”。

(6) 超媒体检索 超媒体检索是对超文本检索的补充。其存储对象超出了文本范畴,融入了静态或动态的图形、图像、声音等多种媒体信息。信息的存储结构从单维发展到多维,存储空间亦在不断地扩大。

### 三、信息检索语言

信息检索语言是为建立信息检索系统而创建的专门用来描述文献特征(内容特征或外表特征)和表达检索提问的一种人工语言,又称为信息存储与检索语言、标引语言、索引语言等。它的主要功能是:① 简单明了而又较为专指地描述信息的主题概念;② 容易地将概念进行系统排列;③ 便于检索时将标引用语与检索用语进行相符性比较。因此,信息检索语言不但需排除一词多义、多词一义和词义含糊的现象,而且要显示出概念间的相互关系,这也是信息检索语言规范化的主要内容。

信息检索语言是决定检索系统中大量信息排检序列的关键。它可以是一系列概括信息内容的概念及其相互关系的标识系统,如分类号码;也可以是自然语言中选择出来并加以规范化的一套词汇,如主题词表。世界上有许多种信息检索语言,我们常用的有:

#### 1. 分类检索语言

分类检索语言是以学科分类为基础,结合信息内容特征的一种直接体现知识分类概念的检索语言。其采用概念逻辑分类的一般规则进行层层划分,构成具有

上位类和下位类之间隶属关系、同位类之间并列关系的概念等级体系。例如：

R5 内科学

R51 传染病

R52 结核病

R53 寄生虫病

R54 心脏、血管(循环系统)疾病

R541 心脏疾病

.1 先天性心脏血管病

.2 风湿性心脏病

.3 高血压性心脏病

.4 冠状动脉(粥样)硬化性心脏病(冠心病)

分类检索语言的“语词”就是它的类目及相应的分类号,分类号主要用于明确各类目之间的先后顺序。如上例的分类号排序是 R5, R51, R52, R53, R54, R541, R541.1, R541.2, R541.3, R541.4……

分类检索语言既可以用于期刊论文的分类,也可以用于图书等其他文献信息的分类。国内外有多种广泛使用的著名分类检索语言,如《美国国会图书馆图书分类法》(Library of Congress Classification, LC)、《国际十进分类法》(Universal Decimal Classification, UDC)、《杜威十进分类法》(Dewey Decimal Classification and Relative Index, DC 或 DDC)、《中国图书馆分类法》(中图法),后者是我国使用最普遍的一种分类检索语言。

《中国图书馆分类法》共分 22 个基本大类(见表 1-1)，“R 医药、卫生”类下分 17 个二级类目(见表 1-2)。

表 1-1 《中国图书馆分类法》基本大类

A 马克思主义、列宁主义、毛泽东思想、邓小平理论	N 自然科学总论
B 哲学、宗教	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、法律	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、安全科学
K 历史、地理	Z 综合性图书

表 1-2 “R 医药、卫生”的二级类目

分类号	类 目	分类号	类 目
R1	预防医学、卫生学	R74	神经病学与精神病学
R2	中国医学	R75	皮肤病学与性病学
R3	基础医学	R76	耳鼻咽喉科学
R4	临床医学	R77	眼科学
R5	内科学	R78	口腔科学
R6	外科学	R79	外国民族医学
R71	妇产科学	R8	特种医学
R72	儿科学	R9	药学
R73	肿瘤学		

## 2. 主题检索语言

主题检索语言是用表达文献主题内容的词语作为标识的信息检索语言。应用较多的是主题词和关键词。

(1) 主题词法 主题词(Subject Heading)又称叙词(Discriptor),是以规范化为基础,以揭示事物对象及其特征为出发点的信息检索语言。最具代表性的主题词法是美国国立医学图书馆(National Library of Medicine, NLM)的《医学主题词表》(Medical Subject Headings, MeSH), MeSH 也是医学领域内使用最多的一种主题检索语言。MeSH 用于标引和揭示医学文献的主题内容,对于提高医学信息检索的准确率具有十分重要的意义(详见第三章第一节)。

随着 Internet 的不断发展、人类信息需求的日益增长,人们在日常的信息检索过程中,越来越重视的是事物的概念和语义,而不容易理解数据库系统的特定句法。所以,自然语言的检索更容易为人们所接受。但使用传统的自然语言检索,由于检索词与著者使用的文本词不统一,容易造成漏检和误检。因此,在计算机信息检索数据库中,出现了检索词自动转换系统、智能检索系统等,这些系统是将用户输入的概念和语义自动转换成满足相对查全和查准的数据库系统语言进行检索。如美国 NLM 自 1986 年起研究和开发的一体化医学语言系统(Unified Medical Language System, UMLS),就是在 MeSH 基础上,应用先进的计算机信息技术建立的一个全新的生物医学信息检索语言的集成系统和机读信息资源指南系统。可用于跨数据库的词汇转换,具有一定的数据库集成检索功能和自然语言词语转换等智能检索功能。

UMLS 通过将大量的检索词(包括规范词和自由词)累积输入系统中,进行检索词自动转换处理,使用户能够不必考虑检索词的规范性或知识分类属性,不受人工语言和自然语言的束缚与限制,更自由地在电子病案、文献数据库、图像数据库、



专家系统等各种信息资源库中检索和获取特定的信息。UMLS 包括 4 个部分:

- 超级叙词表(Metathesaurus):有人译为元辞典,是 UMLS 的核心部分,2001 年版收录了 80 万个概念共 190 万个词汇。这些概念和词汇来自包括 MeSH 在内的 60 多个生物医学词表、分类表、术语表、专家系统等。其目的是要整合各来源词表中的生物医学概念、术语、词汇及其等级范畴,构建一个集成系统,来解决因为各系统的差异性和信息资源的分散性所造成的检索困难。

- 语义网络(Semantic Network):语义网络把概念进行分型或分类,构建概念之间的相互关系,并提供相关信息的获取。例如,查找某病毒的概念时,不但可以获取该病毒的概念和信息,还可以找到该病毒可能引起的疾病或综合征的相关概念和信息。

- 信息资源图(Information Source Picture):是各种生物医学数据库的信息资源集合图,图中描述了各信息资源的范围、定位、词汇、句法和访问条件。其信息资源既可以供人类阅读,也可以被机器处理。

- 专家词典(Specialist Lexicon):是为超级叙词表中的许多术语提供各种构成词的句法信息,也包括没有出现在超级叙词表中的英语单词,如动词。

(2) 关键词法 关键词(Keyword)是指出现在文献的题名、摘要或全文中,能够反映文献主题内容的专业名词或术语。关键词直接取自原文,不作规范化处理,可以提供更多的检索入口,适合计算机系统自动编制索引的需要。但由于词语没有规范化,不能进行选择和控制,容易造成漏检和误检。

分类检索语言和主题检索语言是典型的基于文献内容特征的检索语言,也是信息检索领域使用最多的检索语言。除此之外,人们在日常的信息检索过程中,经常用到的检索语言还有代码检索语言,如美国《化学文摘》(CA)中的分子式检索语言,依据文献外表特征而设计的各种检索语言,如题名、著者、文献序号、引文检索语言等。

#### 四、信息检索工具

信息检索工具是将大量分散无序的信息经过搜集、加工和整理,按照一定的规则和方法进行组织和系统排列,用以报道、存储和查找信息的工具。主要由使用说明、正文(或数据库)、辅助索引和附录四个部分组成。正文(或数据库)部分是信息检索工具的主体部分,它将收入检索工具的每篇文献著录成为题录、文摘或数据库的记录,并将它们按照一定的方式(如学科分类、主题或序号)组织排列成一个有序结合体。辅助索引通常有多种,如主题索引、著者索引、关键词索引、药物名称索引、分子式索引等等,以提供更多的与主体部分不同的检索途径。常见的检索工具有如下类型:

##### 1. 手工检索工具

手工检索工具是由检索者直接利用查找的一类工具,如各种手工检索目录卡、