

医学文献检索与利用

梅 谊  主编



苏州大学出版社
Soochow University Press

医学文献检索与利用

主 编 梅 谊

副 主 编 邱 悦 缪幽竹

图书在版编目(CIP)数据

医学文献检索与利用/梅谊主编. —苏州: 苏州
大学出版社, 2011. 8
ISBN 978-7-81137-747-7

I. ①医… II. ①梅… III. ①医学—情报检索—高等
学校—教材 IV. ①G252.7

中国版本图书馆 CIP 数据核字(2011)第 153797 号

书 名: 医学文献检索与利用

作 者: 梅 谊 主编

责任编辑: 廖桂芝

组 稿 人: 陈林华

装帧设计: 刘 俊

出版发行: 苏州大学出版社(Soochow University Press)

社 址: 苏州市十梓街 1 号 邮编: 215006

印 刷: 丹阳市兴华印刷厂

网 址: www.sudapress.com

邮购热线: 0512-67480030

销售热线: 0512-65225020

开 本: 787 mm×1 092 mm 1/16 印张: 16.5 字数: 410 千

版 次: 2011 年 8 月第 1 版

印 次: 2011 年 8 月第 1 次印刷

书 号: ISBN 978-7-81137-747-7

定 价: 30.00 元

凡购本社图书发现印装错误,请与本社联系调换。服务热线: 0512-65225020

前言

21 世纪是以现代科学技术为核心、以知识创新和技术创新为特征的信息社会。这个时代使人们深切地感受到:知识就是力量,信息就是财富。随着科学技术的飞速发展,信息技术和手段也日新月异,新经验、新知识不断产生,文献数量也不断增加,网络和电子文献资源蓬勃发展。信息化浪潮一方面为我们带来了丰富的文献信息资源,另一方面也为文献信息检索和利用带来了新的挑战。如何及时有效地进行文献检索,以及利用这些记载人类科技成果的文献促进科学研究的进一步发展,已成为科技工作者日益关注的问题。

文献检索是关于如何查阅、获取和利用有效文献的、具有很强的实践性和应用性的一门方法课程。自 1984 年我国教育部颁发关于在高等学校开设《文献检索与利用》课程的文件以来,全国各类高校已将这门课程纳入必修课程体系。医学文献是科技文献的重要组成部分,对于研究人类生命的医学工作者来说,掌握和提高文献检索的方法和技能,可有效地利用医学文献,促进自身的学习、工作和研究,进一步提高自身的科研水平,成为一名具有较强的情报意识和较高的信息素养,善于在现代信息社会中摄取各类信息,善于终身学习,勇于创新的医学工作者。

本书共分 8 章 37 节,内容包括医学文献概论、医学文献检索基础、常用中外文医学文献检索工具(包括 CBM、IM、MEDLINE、EM、BA、CA 等数据库)、药学文献检索(包括中外药典和药物索引等)、常用中外文全文数据库、网络医学资源利用、循证医学、医学论文写作等。本书在原教材《实用现代医学文献检索》的基础上,根据各数据库的改版情况,经过反复修改、补充,并在借鉴国内已经出版的一些比较好的教材的基础上,结合苏州大学实际教学经验及日常检索工作体会编写而成。本书主要介绍常用的医学文献网络数据库的检索方法及 Internet 医学信息资源的利用。对重点内容采用大量的原始图片示例,具有真实、直观、形象、易教、易学、易懂的特点。内容新(大部分数据资料统计截止到 2011 年 4 月),实用性强,图文并茂,便于自学。

本书适用于生物医学专业的研究生、长学制本科生的医学文献检索教学,也可作为临床医生、医学科研和教学人员继续教育的参考用书。本书由长期从事医学文献检索教学、实践经验丰富的教师编写。由于水平、经验所限,书中难免存在缺点和不足,敬请同行专家和广大师生批评指正。

本书在编写过程中参考了大量的重要著作,对提高本书的质量起了十分重要的作用。苏州大学图书馆的有关领导、同仁及苏州大学出版社对本书的编辑和出版给予了大力支持和帮助,在此谨向他们表示衷心的感谢。

梅 谊

2011 年 5 月 1 日 于苏大炳麟图书馆



Contents 目录

第一章 医学文献概论	(1)
第一节 文献、信息、情报和知识	(1)
第二节 文献的类型和级别	(3)
第三节 医学文献的现状和发展趋势	(5)
思考题	(7)
第二章 医学文献检索基础	(8)
第一节 文献检索与检索原理	(8)
第二节 检索语言	(9)
第三节 传统的手工检索工具	(12)
第四节 计算机检索及其数据库基本知识	(14)
第五节 文献检索方法与检索途径、步骤	(18)
第六节 文献检索效果评估	(22)
思考题	(23)
第三章 常用中外文医学文献检索工具	(24)
第一节 美国《医学索引》与医学主题词表(MeSH)	(24)
第二节 中国生物医学文献服务系统(SinoMed)	(33)
检索实习题	(59)
第三节 MEDLINE 数据库(OVID)	(60)
检索实习题	(74)
第四节 美国生物学文摘数据库(OVID-BP)	(97)
检索实习题	(105)
第五节 荷兰医学文摘数据库(EMbase)	(108)
第六节 美国化学文摘数据库(SciFinder)	(118)
检索实习题	(129)
第七节 Web of Science	(130)



检索实习题	(148)
第四章 药学文献检索	(149)
第一节 《中国药学文摘》及其数据库	(149)
第二节 美国《国际药物文摘》(IPA)及其数据库	(154)
第三节 药典和药物索引	(159)
第五章 常用中外文全文数据库	(168)
第一节 全文数据库概述	(168)
第二节 中文全文数据库	(169)
第三节 外文全文数据库	(186)
第四节 超星数字图书馆	(199)
检索实习题	(202)
第六章 网络医学资源利用	(203)
第一节 网络基础知识	(203)
第二节 搜索引擎	(205)
第三节 网络医学资源	(208)
第四节 PubMed	(212)
检索实习题	(225)
第五节 Highwire Press	(225)
检索实习题	(231)
第七章 循证医学	(232)
第一节 循证医学概述	(232)
第二节 循证医学资源及其检索	(234)
第八章 医学论文写作	(241)
第一节 医学论文的写作要求	(241)
第二节 医学论文写作的三大要素	(243)
第三节 医学论文的种类	(243)
第四节 医学论文的体裁	(245)
第五节 医学论文写作的格式要求	(246)
第六节 医学论文写作的步骤	(252)
第七节 医学论文写作的常见问题	(255)
参考文献	(257)



第一章

医学文献概论

第一节 文献、信息、情报和知识

一、文献(literature)

凡是用文字、图形、符号、声频、视频等手段记录在各种载体上的人类知识都可称为文献,简言之,文献是“记录有知识的一切载体”(《中华人民共和国国家标准:文献著录总则》,GB3792.1—83)。文献是由三个要素构成的:(1)有一定的知识内容;(2)要有用以保存和传递知识的记录方式,如文字、图形等;(3)要有记录知识的载体,如纸张、光盘等。文献是社会发展的产物,它记录了人类历史长河中社会活动、生产活动、科技发展所达到的成就和水平,凝聚着数千年来人类辛勤劳动的智慧,是社会精神财富的重要组成部分。科技文献是文献的重要组成部分,而医学文献又属于科技文献范畴。它记录了千千万万医学工作者在研究人类生命过程及同疾病作斗争过程中形成的科学知识,包括研究人类生命活动与外界环境的关系,研究人类疾病的发生、发展及其防治、消灭的规律,以及增进健康、延长寿命和提高生活质量的有效措施等。文献是一种很宝贵的精神财富。

二、信息(information)

自20世纪50年代信息概念被正式提出以来,信息得到了广泛而深入的研究,但由于信息涉及的领域广、内容丰富,至今都没有形成统一的认识。已有的较具代表性的信息定义如下:(1)信息是客观世界中各种事物变化与特征的最新反映,是客观事物状态经过传递后的再现;(2)信息是实现事物间根据某种自然的规律和人为的约定建立联系的一种形式,是被表现出来的事物增添了确定性或被取消了的不确定性;(3)信息是经过传递为接收者所理解,并对解决面临问题有用的、预先不知道的新报道和新知识;(4)信息是物质存在的一种形式,它是以物质的属性和运动状态为内容,并且总是借助于一定的物质载体传输和储存的。

目前较为普通的信息定义为:信息是通过一定的物质载体形式反映出来的事物存在的状态、运动形式、运动规律及其相互作用的表征。信息普遍存在于整个自然界,无处不在,无时不有,其外延是相当广泛的,宇宙万物所产生和彼此交流的内容都可成为信息,如



病人体温的升高与下降是其体征信息;天空乌云密布是暴风雨来临的预示信息;敌军布防的情报是战斗决策信息;DNA 的密码是生命繁衍所依靠的遗传信息;医学成果的论文是有关学术知识的文献信息。信息、能源和材料已构成现代社会文明的三大支柱。信息本身不能独立存在,必须依附于某些物质载体,因此产生了信源、信宿、信道三大概念。信源是信息的发源地或源泉;信宿是信息的归宿处,是信息的接受体;信道即信息传递的通道,是信源和信宿之间的流通渠道。信息具有三个最基本的特性,即差异性、特征性、可传递性。(1) 差异性:信息的意义就在于反映差异,没有差异就不称其为信息。(2) 特征性:信息要反映客观事物在时间、空间上的不同状态。不同事物具有不同的状态,同一事物在不同的时间和空间上也有不同的状态呈现,这就是事物的状态特征,信息则是再现事物状态特征的客观描述。(3) 可传递性:信息是可以传递的,也只有通过传递,才能实现其价值。除了上述三个基本特性之外,信息还具有以下一些物质特性:(1) 普遍性与可识别性:信息无处不在,并能通过人的感官和各种探测设备和手段加以识别。(2) 可存储性与可处理性:借助于纸、磁、光、电等物质载体,可对信息加以保存,同时还可对其进行加工、整理等处理。(3) 时效性与可共享性:信息的价值往往体现在某一个时间段内,具有较强的时效性。信息应属于全人类的财富,不由专人占有,实现信息共享是充分发挥信息价值的最佳途径。(4) 动态性与再生性:信息的内容不是一成不变的,随着时空的改变呈现出动态性变化,并可通过语言、文字、音像等形式再生,呈几何增长。

关于信息素养:信息素养是科研人员对信息社会中人的信息行为能力和思维方式的整体描述,包括信息发现、评价、利用、交流等各种能力,是一种综合能力和素质。信息素养能够使科研人员认识到信息的价值,并能利用信息在生活、工作、学习中作出正确选择,具有信息素养的人能对不同形式、不同内容和不同来源的信息进行有效的收集、评价、组织和综合利用。

三、情报(information)

情报是关于某种情况的消息和报告,是被传递的知识或事实。著名科学家钱学森认为:所谓情报,就是为了解决一个特定的问题所需要的知识,是激活了、活化了的的知识,以及这种知识的及时性和针对性。情报的基本属性包括知识性、传递性和效用性。(1) 知识性:情报来源于知识,任何情报都具有一定的知识。(2) 传递性:知识和信息要转化为情报,必须经过传递,并为用户接受和利用。(3) 效用性:人们创造情报、传递情报的目的就在于充分利用情报,提高其效用性。情报的效用性表现为启迪思维,增进见识,改变知识结构,提高认识能力,帮助人们改造世界,发挥其使用价值、经济价值和社会价值。

关于情报意识:情报意识是指人们对情报在医学教学、科研和临床工作中的地位、作用、价值等的认知程度,对医学情报重要性的认识程度和需求的迫切程度以及捕捉、分析、判断和吸收情报信息的自觉程度。

关于情报能力:包括情报的收集、处理、利用、交流能力。





四、知识(knowledge)

知识是人类对客观世界的正确认识以及对社会实践及生产实践的经验总结,是人类的主观世界对客观世界的概括和反映,人类通过信息来认识世界和改造世界,并在这个过程中又不断地将感性认识或经验总结成知识,或根据所获得的信息上升为知识。例如,1862年俄国生理学家谢切诺夫发现刺激青蛙的丘脑能抑制屈肌反射以及脊髓中的其他反射,其后出版了《大脑反射》一书,他在获得信息的基础上,又将信息组成为知识。知识按其内容可分为自然科学知识和社会科学知识。自然科学知识是人们在改造自然中所获得的知识;社会科学知识是人们在改造社会的实践中所获得的知识。医学知识属于自然科学知识范畴,是人们在长期与疾病作斗争的反复实践过程中所积累的经验结晶。

五、文献、信息、情报和知识四者之间的相互关系

信息、知识、情报和文献之间有着十分密切的关系。信息是事物存在方式和运动状态及其特征的反映;知识是人类对大量信息进行思维、提炼、优化和系统化的结果;情报是被活化了的有用知识;文献则是记录知识的一切载体,它是重要的知识源和情报信息源,也是知识、情报信息存储的重要方式。信息可以成为情报,但是一般要经过选择、综合、研究、分析、加工等过程,也就是要过去粗取精、去伪存真、由此及彼、由表及里的提炼过程。信息是知识的组成部分,但不是全部,只有提高了的、深化了的、系统化的信息才能称得上知识。在知识或信息的海洋里,变化、流动最活跃的那一部分或被激活了的那一部分就是情报。知识、情报、信息的主要部分被包含在文献之中,有人把信息比喻为大海,把知识比喻为大海中的一片水域,而情报则是这片水域中的一个小岛。文献则属于知识的范畴,是知识的一部分。换言之,信息包含着知识,知识包含着文献和情报;信息既包含着知识,又包含着文献和情报。

第二节 文献的类型和级别

一、文献的类型

根据文献的载体形态,可将其分成以下几种类型。

(一) 印刷型(printed form)

印刷型文献即传统的纸质文献,包括石印、油印、铅印、胶印、激光照排的文献,是图书情报机构收藏最多的、使用最广的文献。其特点是符合人类传统的阅读习惯,实用、方便。主要有以下三大类:

1. 图书(book)

图书是现代出版物中最普通的一种,历史久、内容广、数量多。它是系统地论述某一专题比较成熟的、定型的带有总结性的文献。一般都有固定的装帧,通过出版社出版发





行,有国际标准书号(ISBN)。图书可分为两大类:一类是供读者阅读的书籍,如教科书、专著、丛书、多卷书集等;另一类是供读者检索参考用的工具书,如书目、索引、文摘、手册、辞典、年鉴、图谱、百科全书、指南等。

2. 期刊(journal)

期刊是指连续、定期、长期出版的出版物,有固定的刊名,每期版式大致相同,有连续的年、卷、期号,有国际连续出版物号(ISSN)。期刊出版周期短,速度快,内容新颖,信息量大,是主要的信息源,也是文献检索的主要对象。期刊按内容性质的不同可分为:学术性、科普性、商业性、娱乐性等类型。

3. 资料(material)

资料是指图书、期刊以外的出版物(无ISBN号和ISSN号)。如科技报告、政府出版物、专利文献、产品样本、学位论文、技术标准、会议论文等,常为不定期出版,也可能不公开出版。由于此类文献有许多涉及最新研究和技术及国家的法规、标准等信息,故也是科研的重要信息源。

(二) 缩微型(microform)

缩微型文献是指采用照相技术,将文献缩小拍摄在感光胶卷或平片上,通过专门的阅读机进行阅读的文献,包括缩微胶卷、缩微胶片等。其优点是体积小、容量大、易保存、成本低,便于携带、存储。缺点是阅读不方便。

(三) 视听型(audio visual form)

视听型文献亦称声像型,包括唱片、录音带、录像带、电影胶片、幻灯片等。此类文献形象直观,读者容易理解,便于掌握,能反复播放和录制。

(四) 机读型(machine readable form)

机读型文献是指通过计算机存储阅读的文献。它以数字形式将文献存储在磁盘、光盘等介质(如磁盘、磁带、磁卡、光盘)——计算机的外存储器中,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。其优点是存储量大,检索速度快,使用快速、方便。随着计算机及网络技术的发展,此类文献将越来越受到人们的青睐。缺点是费用较大。

(五) 书写型(write form)

近年来,有人提出书写型文献,也就是手稿,虽也属于文献范围,但很少在正规渠道传播。

二、文献的级别

文献根据内容性质、加工深度的不同,可分为以下三个级别:

1. 一次文献(primary literature)

一次文献是未经加工的原始文献,是由作者直接记录的生产、科研成果或报道的新发明、新创造、新技术、新知识、新见解等,具有新颖性、创新性的特点。如期刊论文、专利文献、技术标准、科技报告、会议论文、学位论文等,它是文献检索的主要对象。

2. 二次文献(secondary literature)

二次文献是将分散的无组织的一次文献进行加工、整理、编排(总称为文献标引)的



产物,即检索工具。它是用来查找一次文献的线索,包括目录、索引、文摘等。一般而言,一次文献发表在前,二次文献编制在后。但近年来由于科技参考的需要,人们要求缩短文献出版的时差,因此,有些出版物在发表原文前,先发表题录或文摘,如美国《现期期刊目次》(Current Contents)就是在原始论文尚未出版之前,预先报道了一次文献的线索。所以,一次文献与二次文献的关系正在发生变化。

3. 三次文献(tertiary literature)

三次文献是在利用二次文献的基础上,选用大量一次文献的内容经过综合、分析,并加上作者的评述撰写而成的,包括综述、述评、年鉴、手册、进展、数据手册及指南等,是经作者浓缩、再生的科研文献。

此外,目前还有零次文献的提法。零次文献是指尚未形成文字记载、未出版的口头交流的文献,包括专家们口头交流的经验,科研设想,现场考察所见所闻,某些专门技巧,仪器、设备使用的窍门,医务人员的临床心得,教师的教学经验等,其中有的可能永远不会正式以文字形式报道。

从一次文献到二次文献、三次文献,是一个由分散到集中、由无组织到系统化的过程。对于广大信息用户来说,一次文献是检索的对象,二次文献是检索的工具,而三次文献则是情报调研的结果。各级文献的产生状况见表 1-1。

表 1-1 各级文献的产生状况

	生产者	生产工艺	产品形态
一次文献	科研人员、 临床工作者	科学实验、 临床观察	期刊论文、科研报告、 学位论文、专利等
二次文献	文献标引工作人员	收集、加工、整理	目录、索引、文摘(检索工具)
三次文献	专家学者	综合分析、 浓缩重组	综述、手册、年鉴、百科全书等

第三节 医学文献的现状和发展趋势

现代科学技术正朝着高度综合又高度分化的方向发展,学科越分越细、分支越来越多,学科之间相互交叉、相互渗透,边缘学科、新兴学科不断涌现,知识门类日益增多。作为记载、保存和传播知识的文献出现了极其复杂的局面,主要表现在以下几个方面。

一、数量庞大

近 20 年来科学家们创造的新知识,几乎相当于几千年全人类知识积累的总和。人类知识的迅速增长,导致科技文献数量激增。而医学文献增长之迅速、数量之庞大,居各学科之首。目前,全世界的生物医学期刊已超过 2.1 万种,约占世界期刊的 1/7。《科学引文索引》(SCI)按引文数量所排的前 500 种期刊,其中医学有 176 种,约占 1/3。其他类型



的医学文献,如专利、会议录、学位论文、科技报告、产品说明书等的情况基本类似,可见其数量之庞大。

二、类型复杂

知识载体的多样化,使得文献的类型也十分复杂,除传统的印刷型文献外,视听型、缩微型、机读型等文献,以其出版速度快、存储密度高、形象直观、便于携带等优点而发展迅速,特别是随着多媒体技术的应用,更多的新型载体文献将会大量涌现。

三、文种繁多

全世界出版文献的文种在不断地增加,目前已达到 70 ~ 80 种,科技期刊较常用的文种有 6 ~ 7 种,其中英文占 60%,俄文占 10%,德文占 10%,法文占 7%,日文占 3%,西班牙语占 2%,中文及其他文种占 8%。据统计,美国 MEDLINE 数据库收录文献的语种最多时达 76 种。文献文种的不断扩大,造成了使用文献时的语言隔阂,阻碍了国际间的科学交流和情报传递,也给科研工作者掌握各门学科知识、及时获取最新信息、更新知识增加了难度。

四、重复交叉

随着科学技术的发展,科技活动日益繁多,反映科技进展的形式各异,同一篇论文往往以多种形式出版,一项专利同时可以在几个国家申请,一篇关于免疫学的文献可以从分子生物学、肿瘤学、医学工程等多种杂志上获取,造成大量重复交叉。

五、分布分散

当前科学技术发展的特点是学科越分越细,内容彼此渗透、相互交叉,加上期刊种类越来越多,致使各专业文献发表分散。布拉德福定律表明,一个专业的文献,只有 1/3 刊登在本专业的刊物上,1/3 刊登在综合性刊物上,1/3 刊登在相关学科的刊物上。文献的这种分散现象是现代科学门类之间广泛交叉渗透、分化综合这一特点的必然反映。

六、更新速度加快

随着科学的发展、知识的更新,文献也在老化。科技文献的老化周期,已从 19 世纪的 50 年左右缩短到目前的 5 ~ 10 年。有人认为,当今世界 80% ~ 90% 的科技文献其有效期仅 3 ~ 5 年,医学文献的老化速度更快,其“半衰期”已不足 5 年。

七、交流传播速度快

计算机技术、通讯技术、多媒体技术的广泛普及、应用,为文献的快速传播提供了便捷



的条件。Internet 打破了时空界限,推动了全球信息化。现在世界上发生的重大事件,瞬息即可耳闻目睹,某种疾病治疗方法的突破或特效药物的产生,立刻就会在世界各地传播。随着现代高新技术的发展,文献信息交流的速度会越来越快,文献的寿命越来越短。

八、向数字化、网络化发展

进入数字化时代,文献的出版、传递方式发生了很大的变化,印刷型文献经数字化处理,制成可供计算机阅读的电子出版物,如电子图书、电子杂志、电子地图等,亦可通过因特网进行传递、检索和利用。数字化、网络化的发展,使信息、文献实现了真正意义上的资源共享。



思考题

1. 什么是文献?
2. 文献有哪些类型?各有什么特点?
3. 文献的级别指的是什么?
4. 医学文献的现状如何?



第二章

医学文献检索基础

第一节 文献检索与检索原理

一、文献检索

文献检索指根据课题需要,运用科学的方法,采用专门的工具,从大量文献中尽可能迅速、准确而无重大遗漏地获取所需文献的过程。文献检索按检索内容要求的不同,具体可分为事实检索、数值检索、文献(书目、全文)检索;按检索数据揭示内容程度的不同,可分为题录检索、文摘检索和全文检索;按文献检索的手段不同,又可分为手工检索和计算机检索;按文献检索的检索标记的不同,又可分为分类检索、主题检索、著者检索;等等。文献检索是信息检索的一个重要组成部分,科技人员在进行信息检索的过程中,通常以文献检索为主。

文献检索的作用及意义:对于科研人员来说,学会文献检索有利于不断学习专业知识,拓宽知识面,能够捕捉更多的信息,增强情报意识,立足于社会。学会文献检索,可节省查找文献的时间和费用,提高文献信息的利用率,促进信息资源的开发利用;避免低水平的重复劳动;有利于全面掌握必要的文献信息;能提高和增强信息素质,加速成才。文献检索的意义包含三个方面:一是确定有无类似的研究成果,是否要立项;二是参考其他文献,理清思路,搞清楚要做的研究是什么;三是取长补短,推陈出新,设立新论题,取得新成果。文献检索在课程设计、毕业设计、开题立项、成果查新等方面具有重要作用。

二、检索原理

广义的文献检索包括文献的存储和检索。文献的存储过程,是由文献标引专业人员完成的。他们将分散的文献资料进行搜集、评价和选择,然后对确定收录的文献信息进行著录、标引、编写文摘等工作,即进行文献特征描述、加工,并使其有序化,组织成具有检索价值的数据库系统供检索使用。文献存储结果的表现形式即是检索工具,如题录、文摘、数据库等。文献的检索过程就是用户利用检索工具获取所需文献的过程。用户在分析待检课题内容范围的基础上,将检索需求转换成检索提问标识,利用相应的检索系统,查出所需文献。存储与检索是互逆的过程。存储是检索的基础,检索是存储的目的,存储



的质量高低直接影响到检索的效率;而检索则是存储的反过程,检索的复杂性又促进了存储水平的不断提高。检索效率的高低取决于标引与检索时是否用相同的标识词,标识一致,文献命中率高。这个“标识”,也就是检索语言。可见,检索语言是文献检索的枢纽,是联系标引人员与检索用户之间的桥梁。

第二节 检索语言

检索语言是为建立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。

一、检索语言的分法

检索语言有很多种,按照不同的标准可有不同的划分,归纳起来主要有以下几种分法。

(一) 按检索语言的受控情况划分

1. 规范化语言

规范化语言又称受控语言、人工语言,是采用经过人工控制的规范性的专业名词术语或符号作为检索语言,用来专指某一概念,这些规范化的语言能较好地对同义词、近义词、相关词、多义词及缩略词等概念进行严格的规范化处理。

2. 非规范化语言

非规范化语言又称非受控语言、自然语言,是采用未经人工控制的专业名词术语或符号作为检索语言,通常所说的自由词、关键词就属此类。当某些特定概念无法用规范词准确表达,或新出现的词语还未来得及被规范化时,只能使用非规范词。这类词语有较大弹性和灵活性,检索者可以自拟词语进行检索。在机检中,非规范词的应用比较广泛。但这类语言对一词多义、多词一义的词语,检索就相对困难些。不同的检索语言构成不同的检索标识和索引系统,提供用户不同的检索点和检索途径。

(二) 按检索词组配设置的时间划分

1. 先组式语言

先组式语言是指复杂主题的检索词在检索前已经在检索系统中组配好了的检索语言,如分类语言、主题词语言。

2. 后组式语言

后组式语言是指在检索前检索系统中的检索词是独立的,实施检索后才根据检索需要对检索词进行组配的检索语言,如关键词语言、单元词语言等。

检索的匹配就是通过检索语言的匹配来实现的。检索效果的好坏,在很大程度上取决于所采用的检索语言的质量以及使用是否正确。检索语言除表达事物不同概念之外,有时还要揭示概念之间的相互关系。概念之间的相互关系主要有以下几种:

(1) 等同关系:指两个或两个以上的词所表达的概念完全或基本相同,如同义词、近



义词、全称与简称等。

(2) 从属关系:是指两个概念中,一个概念被包含在另一个概念的外延中,是另一概念的外延部分,也就是上、下位概念的关系,外延较大的是上位概念,较小的是下位概念。如高血压与肾性高血压。

(3) 相关关系:是指概念之间的关系的密切程度,不同于等同关系和从属关系的一种关系。如交叉关系、矛盾关系、对立关系、并列关系等。

二、常用的检索语言

(一) 内容特征检索语言

内容特征检索语言包括能表达文献内容特征的词、代号等。

1. 分类语言

分类法是以学科专业为基础,根据文献的内部或某些外部特征,运用概念划分的原理,将知识门类按照一定的逻辑次序分门别类地排列起来的一个等级体系。分类的原则是从总到分、从一般到具体、从简单到复杂,层层划分,逐级展开。一个上位类往往划分出许多平行的下位类,一个类目向上是隶属关系,向下是派生关系,从而形成一个有序的、直线型的知识门类等级体系。分类表中的类目又称类名,每个类目配一个号码称类号。体系分类法能较好地体现学科的系统性,揭示知识之间的隶属、派生和平行的关系,便于从学科专业的角度查找文献,还便于扩大或缩小检索范围。分类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)等。《中国图书馆分类法》是我国使用最普遍的一种分类检索语言,共分五大部分 22 个基本大类,“R 医药、卫生”大类下共分 19 个二级类目(包括 R-总论和 RZ 地理名称)(图 2-1)。CBM 数据库即按中图法分类检索。

2. 标题词语言

标题词语言是指用来表达文献主题内容的、有实质意义的专业名词术语作为标识的检索语言。应用较多的是主题词和关键词。

(1) 主题词:主题词(subject heading)又称叙述词(descriptor),是用来表达文献主题概念的有实质意义的经过规范化处理的(分析转换)专业名词或词组。最具代表性的主题词表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,对于提高医学文献检索的准确率具有重要的意义(参见本书第三章第一节)。

(2) 关键词:关键词(key word)是指出现在文献的题名、摘要或正文中,有实质意义的能表达文献中心内容的词语,即对揭示和描述文献主题内容而言是重要的、关键性的词语。关键词直接取自原文,一般不作规范化处理,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。但由于关键词没有经过规范化处理,不能进行选择和控制,容易造成漏检和误检。



图 2-1 中国图书馆分类法示意图

主题词与关键词最大的区别就是主题词经过了严格的规范化处理。

主题词是规范化的检索语言,对文献中出现的同义词、近义词、多义词以及同一概念的不同书写形式等进行了严格的控制和规范,使每个主题词都含义明确,以便准确检索,防止误检、漏检。如白介素 2、白细胞介素 2、IL2、IL-2 等表达同一概念的不同书写形式统一规范为“白细胞介素 2”。主题词表是对主题词进行规范化处理的依据,也是文献标引者和检索者共同参照的依据。而关键词是属于自然语言的范畴,未经规范化处理,也不受主题词表的控制。如对于“白细胞介素 2”这一概念可用白介素 2、白细胞介素 2、IL2、IL-2 等不同形式来表达。

因此,为了达到较高的查准率和查全率,如果检索工具提供了主题词这一检索途径的话,就应该选择主题词来进行检索,而不应该选择关键词。

(3) 代码语言:代码语言是依据文献内容特征而设计的检索语言,如美国《化学文摘》(CA)中的分子式、专利号、化学物质登记号、《生物学文摘》(BA)中的生物体名称等。

(二) 外表特征检索语言

外表特征检索语言是用无法表达文献内容特征的词语或代号作为检索语言,如文献题名、著者姓名、期刊名称等。一般采用自然语言直接检索。