

高等学校改革教材

医药电子资源 的检索与利用

主 审 / 王杰贞

主 编 / 马 路 周 纯

副主编 / 黄 芳 黄晴珊

人民卫生出版社

People's Medical Publishing House

高等学校改革教材

医药电子资源的检索与利用

主 审 王杰贞

主 编 马 路 周 纯

副主编 黄 芳 黄晴珊

编 者 (以姓氏笔画为序)

马 路	(首都医科大学)	王 凌	(首都医科大学)
古 明	(中国药科大学)	李 宁	(首都医科大学)
张士靖	(华中科技大学)	张教业	(广州医学院)
周 纯	(中山大学)	陈文凯	(广东医学院)
高 琴	(北京大学)	高 瑜	(郑州大学)
黄 芳	(首都医科大学)	黄晴珊	(中山大学)
龚佳剑	(首都医科大学)	管 进	(广东医学院)
蔡庆芳	(广州医学院)		

人 民 卫 生 出 版 社

图书在版编目 (CIP) 数据

医药电子资源的检索与利用/马路等主编. —北京:
人民卫生出版社, 2006. 2

ISBN 7-117-07434-5

I. 医... II. 马... III. 医药学—因特网—情报检
索 IV. G354.4

中国版本图书馆 CIP 数据核字 (2006) 第 008161 号

医药电子资源的检索与利用

主 编: 马 路 周 纯

出版发行: 人民卫生出版社(中继线 67616688)

地 址: (100078)北京市丰台区方庄芳群园 3 区 3 号楼

网 址: <http://www.pmph.com>

E - mail: pmph@pmph.com

邮购电话: 010-67605754

印 刷: 北京铭成印刷有限公司

经 销: 新华书店

开 本: 787×1092 1/16 印张: 13.75

字 数: 323 千字

版 次: 2006 年 2 月第 1 版 2006 年 2 月第 1 版第 1 次印刷

标准书号: ISBN 7-117-07434-5/R·7435

定 价: 28.00 元

著作权所有, 请勿擅自用本书制作各类出版物, 违者必究

(凡属印装质量问题请与本社销售部联系退换)

前言

医学信息检索课是网络社会和知识经济时代,为适应科技迅猛发展、培养有综合能力人才的需要而开设的一门新型课程。它是以图书馆学、情报学为理论基础,以计算机和网络为手段,侧重培养学生信息意识、获取信息能力的一门科学方法课。

医学信息检索课能够培养学生独立学习、继续学习、接受终身教育的能力,是高校学生充分利用各种信息资源的基础。国家教委明确指出“文检课是培养学生掌握利用文献、情报检索,不断提高自学能力和科研能力的一门科学方法课”。在当今信息社会中,让学生具备获得最新科技信息的能力已成为高等医学教育中不可忽视的重要一环。

近年来,随着信息技术的飞速发展,信息检索系统产品日益丰富,网络信息资源日益膨胀,计算机检索和网络检索已成为文献检索课的“重头戏”,也是文献检索课发展和改革的一个主要方向和趋势。为了适应这个变化,我们根据日常教学的经验,从实用角度出发,从国内外各种电子信息资源中精选出一些优秀的、具有代表性的数据库,通过对医药学信息资源内容和使用方法的介绍,帮助学生、教学科研人员以及临床医务人员掌握检索和利用医药学信息资源,合理合法地检索到自己所需要医药学信息的知识和技能,从而全面提高信息素质。

本书主要适用于医药卫生专业的本科生、研究生。另外,也可作为临床医生,医学、药学科研和教学人员继续教育的参考用书。相对于目前国内同类教材,本教材具有以下特点:

1. 不仅侧重于教授获取信息的手段和方法,同时注意信息素质的教育,如信息道德素质的教育。
2. 讲授的内容不求全,但求实用。目前国内引进和生产的数据库越来越多,每一个数据库都介绍是没有必要的,本书选择了国内外最有代表性的数据库进行了详细的介绍,目的在于培养学生触类旁通的终生学习的能力。
3. 每种资源都精选实例进行分析,并有练习题帮助学生掌握检索技巧。

4. 特意安排一章内容详细介绍如何制定和随时调整自己的检索策略，具有很强的实际指导意义。

5. 任何全文数据库都不可能涵盖所有文献的内容，本书的最后一章重点就在于教会学生在掌握文献线索的时候如何获取全文信息。

在本书的编写过程中，我们参考了大量有关的专著和论文，在此谨向这些作者表示衷心的感谢！同时也感谢首都医科大学学校领导和各相关部门对本书的大力支持！

尽管本书的编委大多具有高级职称或硕士学位，但平均年龄比较小，编写水平有限，加之时间仓促，难免会有疏漏和错误之处，恳请广大读者和同行批评指正！

马 路

2006 年 2 月于北京

目 录

第一章 医药电子资源的检索与利用概论	1
引言	1
第一节 医药电子资源的概念与类型	2
第二节 医药电子资源的检索	5
第三节 医药电子资源的检索途径、检索技术与检索步骤	8
第四节 如何合理使用网络	14
第二章 中文参考数据库	17
第一节 中国生物医学文献数据库	17
第二节 中文生物医学期刊文献数据库	29
第三章 外文参考数据库	36
第一节 MEDLINE 光盘数据库	36
第二节 PubMed 数据库	46
第三节 BIOSIS Previews 数据库	61
第四节 EMBASE.com 数据库	69
第五节 美国《化学文摘》数据库	75
第六节 《国际药学文摘》数据库	86
第四章 中文全文数据库	92
第一节 CNKI 中国期刊网	92
第二节 维普中文科技期刊数据库	96
第三节 万方数据资源系统	103
第五章 外文全文数据库	107
第一节 OVID 数据库	107
第二节 SpringerLink 期刊全文数据库	111
第三节 Elsevier Science 期刊全文数据库	114
第四节 EBSCOhost 期刊全文数据库	117
第五节 ProQuest Medical Library 全文医学期刊数据库	122
第六节 Wiley InterScience	124
第七节 OCLC FirstSearch	126
第六章 引文检索	130
第一节 Web of Science	130

第二节	期刊引用报告·····	138
第三节	中国科学引文数据库·····	139
第四节	中国生物医学期刊引文数据库·····	141
第五节	其他数据库的引文检索功能·····	142
第七章	会议论文与学位论文·····	144
第一节	科学技术会议录索引·····	144
第二节	中国医学学术会议论文数据库·····	148
第三节	ProQuest 博硕士论文数据库·····	152
第四节	中国学位论文全文数据库·····	155
第五节	中国学术会议论文全文数据库·····	157
第六节	中国优秀博硕士学位论文全文数据库·····	158
第八章	网络资源·····	159
第一节	医学搜索引擎·····	159
第二节	常用医学网站·····	164
第三节	网络专类(专题)信息检索·····	166
第九章	网络电子出版物·····	177
第一节	电子图书资源·····	177
第二节	网络参考工具书·····	181
第三节	免费网络期刊·····	186
第四节	网络报纸·····	190
第十章	检索实例分析与检索策略调整·····	193
第十一章	索取原文·····	203
第一节	馆藏联机目录的使用·····	203
第二节	全国期刊联合目录·····	205
第三节	国家科技文献信息网络服务系统·····	206
第四节	CALIS 西文期刊目次数据库·····	209
参考文献	·····	211

第一章 医药电子资源的检索与利用概论

引言

二十世纪下半期,世界经济高速增长,信息资源发挥了极为重要的作用。其中医药学信息资源占全部科技信息的1/4左右,其迅猛的发展对医学、药学、生物学及相关学科的发展起了巨大的推动作用。医药学信息资源发展的特点是:

1. 数量巨大,增长迅速,数量居各学科之首 近年来,医学及生物科学信息增长十分迅速。以占全球出版物总量70%的期刊为例,据统计,全球期刊总数14万余种,其中生物医学期刊超过21000种,占1/7强。《科学引文索引》按引文数量排列的前500种期刊中,生物医学期刊有176种,约占1/3。

2. 内容广泛、分散,各学科相互渗透 由于新的信息载体和传播手段不断出现,各学科相互交叉、彼此渗透加快,信息重复出现的情况越来越多。同一篇文章,可以用不同的形式、不同文字、在不同范围传播和发表。据调查,美国武装部队技术情报局60%的科技报告、美国科学基金会95%的技术情报信息,既出单行本又在期刊上发表,还有可能在网上传播,学术会议文献则90%以上在刊物发表。信息内容的重复,是信息数量剧增的重要原因之一。

各学科发展的相互渗透,导致信息内容呈分散趋势。一篇专题论文涉及几个专业,一种专业期刊所报道的内容又包含了多个学科。通过对专业信息分布状况的分析,约1/3的论文发表在本专业期刊上,1/3的论文发表在相关专业期刊上,另有1/3的论文发表在与本专业无关的期刊上。如体育刊物上发表运动医学论文、音乐杂志发表耳鼻喉科论文都不足为奇。信息内容的重复、交叉和渗透,是现代科学技术广泛融合、渗透的结果,给信息服务工作提出了新的挑战。

3. 信息更新周期缩短,交流传播速度加快 现代社会新理论、新技术、新方法、新成果,不断取代原有的理论、技术、方法和成果。同样,反映科学技术水平和成果的信息的更新速度越来越快,更新周期越来越短。科学信息的更新周期,已经从十九世纪的50年左右缩短到目前的5~10年。某些新技术,如信息技术、微电子技术、无线通信技术等,更新周期更短,甚至以月计。信息内容淘汰速度加快,更新周期缩短,对信息的传播手段提出了新要求。目前,由于网络技术的飞速发展,信息传播交流速度日益加快,如PUBMED、中国期刊全文数据库等网络数据库都是每日更新数据,从这些数据库获取信息已经没有时间和地域的差别。反之,信息传播速度的加快,对信息更新周期的缩短又起到了催化作用。

4. 信息由传统的纸质印刷型迅速向电子化、网络化、数字化方向发展 随着新技术、新材料的发展,相继出现了以感光材料为载体的缩微型文献,以录音、录像带等磁

性材料为载体的视听型文献,以数字化为特征的电子型文献,这一切带来了信息领域的深刻变化。信息资源的发展方向就是数字图书馆。

1992年美国初次提出“数字图书馆”一词,经过十余年发展,该术语已为世界各国所接受,但普遍使用的“数字图书馆”的定义尚未确定。据美国学者统计,1999年数字图书馆相关定义已达65种,上海图书馆2000年也收集到近百种数字图书馆定义。有学者将这一现象概括为“有多少数字图书馆项目就有多少种定义”。定义分散加上近似概念的参与,如数字对象图书馆、多媒体图书馆、电子图书馆、虚拟图书馆、信息高速公路、国家信息基础设施等,使数字图书馆成为二十世纪末不同学科共同关注的热点问题。数字图书馆联盟(DLF)有关数字图书馆的定义是“数字图书馆是对信息进行收集、整理、保管和使用并保证数字信息的一致性,从而能够方便、经济地为用户服务的机构”。其中所包含的突出特征是将图书馆专业人员引入系统建设过程中,从而为图书馆员提供了新的发展空间。Rossatkinson 1996年将数字图书馆描述为“信息控制区域”,在该区域,通过专家选择的高质量信息,成为“资源、浏览和查询机制、分布式网络化环境、信息服务等多种资源和功能的集合体”。“数字图书馆不是单一的实体,而是技术与资源的互联,是为终端用户提供透明式服务、全球化存取的集合体”。

数字图书馆的建设将促进我国信息技术的发展,同时带动与之相关的计算机技术、网络技术、通讯技术和多媒体技术等各项高新技术的迅速发展。这些高新技术迅速转化为现实生产力,将对我国知识创新体系的建立起到极大的促进作用。

以上医药学信息发展的特点与变化都为人们获取和使用带来了新的问题。如何检索和利用医药学信息资源,如何合理合法地检索到自己所需要的医药学信息成为用户信息素质培养的重要问题。本书针对上述问题,通过对医药学信息资源内容和使用方法的介绍,帮助用户解决好这些问题,提高用户的信息检索能力,从而全面提高信息素质。

第一节 医药电子资源的概念与类型

一、电子资源的概念与特性

电子资源(electronic resources),也叫数字资源,指一切以数字形式生产和发行的信息资源。所谓数字形式,是指能被计算机识别的、以不同序列的“0”和“1”构成的形式。电子资源的信息,包括文字、图片、声音、动态图像等,都是以数字代码方式存储在硬盘、磁带、磁盘、光盘等介质上,通过计算机输出设备和网络传出去,最终显示在用户的计算机终端上。医药电子资源是指包含医药学相关知识的电子资源。

随着因特网的发展,利用网络传递的电子资源的数量每年都以几何倍数增长,我们把这一类资源统称为网络电子资源(network electronic resources)。除此之外,目前仍然存在着仅在本地计算机上使用、没有通过网络传递的资源,如光盘数据库和机读磁带数据库等,也属于电子资源范畴。

电子资源的特性为:

1. 易获取性 电子资源借助计算机与网络,向用户提供检索和传递信息的服务,因特网穿越时空的优势使用户在任何时候、任何地点均可通过网络获取电子资源。电子

资源的出现使图书馆的服务在时间和空间上得以延伸,使信息的获取更为便捷。

2. 共享性 与传统文献相比,电子资源使多个用户可在同一时刻下共享同一信息源。传统文献,如一册书或一本杂志,同一时间只能供一个读者阅读。而电子资源则不同,同一个数据库可被多个用户同时检索,同一篇资料可供多个用户同时下载、浏览,体现了电子资源特有的实时共享性,提高了资源的利用率。

3. 信息更新快 在学科发展突飞猛进、知识日新月异的时代,信息的时效性显得尤为重要。在减少了印刷型文献印刷、装订、运输等繁冗环节之后,网上出版物和数据库通常以日或周为更新周期,其高效的动态更新和快捷的实时传递,在充分发挥信息的时效价值方面有着传统文献无可比拟的优势。

4. 形式多样性 多媒体技术在电子资源中的利用,使文字、图形、声音、动画、三维图像等相结合,提供了丰富多彩、生动逼真的信息,使用户能够闻其声、见其形,有身临其境之感。与此同时,超文本技术的运用改变了传统信息的线性组织方法,增强了人们对知识的表达能力,大大拓展了信息的获取和传播范围,使用户更容易理解和接受。

5. 知识的类聚性 受手工检索工具揭示文献的不充分和检索手段繁琐的限制,传统文献中大量有用知识被隐藏。而在网络环境下,依靠强大的检索软件可以使不同类型的相关信息聚集并有序化。目前网络数据库有大型化趋势,同一公司制作的多种数据库产品的检索界面趋向统一,并可实现跨库检索,对检索的完整性提供了可靠保障。除各数据库之间的链接外,数据库还可与图书馆书目查询系统进行链接,这些链接方式可以帮助用户从书目、题录、文摘等二次文献直接查阅到全文。

二、电子资源的类型

医药电子资源的范围非常广泛,其类型多种多样,划分标准也有很多种。

1. 按照电子信息资源的载体形态划分 电子资源按其载体形态划分主要有两大类:一是以光盘、磁盘、磁带等脱机载体出版的有形电子资源;二是在网络上出版的网上电子资源。数据库光盘和多媒体光盘曾经是图书馆电子文献的主要组成部分,但随着因特网的日益普及和飞速发展,网上资源的优势日趋显著。与光盘相比,网上资源不需要光盘塔等设备的投入和维护,内容更新周期更快。因此,网上资源将成为图书馆电子资源的主体,光盘只是作为数据备份的辅助工具,这是一个不可逆转的趋势。

2. 按照电子信息资源的性质和功能划分

一次文献:即原始文献,指作者以生产或科研成果为依据而创作的原始信息、成果、过程以及对其进行分析、综合、总结的信息资源,如事实数据库、全文数据库、电子期刊、电子图书、发布一次文献的学术网站等。包含的内容有:专著、期刊论文、研究报告、专利文献、学位论文、会议文献等。用户可以从一次文献中直接获得所需要的原始信息。

二次文献:指对一次文献进行加工、整理、排序,便于利用一次文献的信息资源,如参考数据库、网络资源学科导航、搜索引擎、分类指南等。二次文献把大量分散的一次文献按照学科或者主题集中起来,著录一次文献的内容和外在特征,组织成相关信息的集合,向用户报道原始信息。它是一种有效的检索工具,为用户提供查找信息的线

索,其形式包括:目录、索引、文摘等。

三次文献:指对二次文献进行综合分析、加工、整理的信息资源,如专门用于检索搜索引擎的搜索工具,被称为“搜索引擎之搜索引擎”(search engine of search engine),即“元搜索引擎”(metasearch engine)。一个典型的元搜索引擎是1995年华盛顿大学硕士生 Eric Selberg 和 Oren Etzioni 推出的第一个元搜索引擎——Metacrawler。在这里,“元”(meta)为“总的”、“超越”之意,元搜索引擎就是对多个独立搜索引擎的整合、调用、控制和优化利用。相对元搜索引擎,可被利用的独立搜索引擎称为“源搜索引擎”(source engine),或“搜索资源”(searching resources)。整合、调用、控制和优化利用源搜索引擎的技术,称为“元搜索技术”(meta-searching technique),元搜索技术是元搜索引擎的核心。当用户进行检索时,反映出来的结果是各搜索引擎的检索结果。

3. 按照电子信息资源的发布形式划分

参考数据库(reference database):指反映各种数据、信息或知识的原始来源和属性的数据库。数据库由记录组成,通过对数据、信息或知识进行加工和过滤,便形成了记录。它主要包括书目数据库、文摘数据库和索引数据库。书目数据库是针对图书、期刊等出版物的报道和揭示,如各图书馆的馆藏机读目录数据库;文摘和索引数据库是针对期刊论文、会议论文、专利文献、学位论文等进行内容和属性的认识与加工,如中国生物医学文献数据库(CBMDISC)、中文生物医学期刊文献数据库(CMCC)、MEDLINE等数据库。

全文数据库(full-text database):即收录有原始文献全文的数据库,包括期刊论文、会议论文、专利文献、学位论文、政府出版物、研究报告、法律条文和案例、商业信息等。如中国期刊全文数据库、中文科技期刊数据库等。

事实数据库(factual database):指包含大量数据、事实,直接提供原始资料的数据库,又分为数值数据库、指南数据库、术语数据库等,相当于印刷型文献中的字典、辞典、手册、年鉴、百科全书、组织机构指南、公式与数表等。例如 MICROMEDEX 数据库、电子版百科全书、网络辞典等。

电子期刊(electronic journal):又名电子杂志(electronic magazine)或数字化期刊(digitized periodical),是具有连续出版物的一般特征,以数字化形式存在,并且仅能通过电子媒介获取的连续出版物。经过20年的发展,电子期刊已从最初的软盘期刊,第二代的CD-ROM期刊、联机期刊,发展到现在第三代的网络化电子期刊。如Springer电子期刊。

电子图书(electronic book):是相对传统的纸质图书而言的,数字化的、以电子文件形式存储在各种磁或电子介质中的图书,如方正Apabi、超星、书生、中国数图有限公司网上图书馆、OCLC的NetLibrary等。

电子报纸(electronic newspaper):是多媒体技术、网络技术和通信技术的产物,不同于某些单纯使用了电子排版技术的报纸或是某些以光盘为载体发行的报纸。真正的电子报纸完全不利用纸张,它是将电子技术应用到出版、发行、利用的全过程。随着科学技术的不断完善,电子纸时代即将到来。电子纸的原理很简单,只是两层用特定的化学物质做透明涂层的薄膜,上面的化学成分作为连续电极。薄膜之间夹有无数黑色、白色

的化学微粒。在薄膜带负电的部分，因吸附带正电的黑粒而呈现黑色；在薄膜带正电的部分，因吸附带负电的氟化碳而呈现白色。只要你拥有一个报社发放的电子报纸阅读器，每天的报纸内容就会以无线的方式发送到你的阅读器中。

搜索引擎/分类指南：搜索引擎通常是指收集了因特网上几千万到几十亿个网页并对网页中的每一个词（即关键词）进行索引，建立索引数据库的全文搜索引擎。当用户查找某个关键词的时候，所有包含了该关键词的网页都将作为搜索结果。在经过复杂的算法进行排序后，这些结果将按照一定的顺序排列。现在的搜索引擎已普遍使用超链分析技术，除了分析索引网页本身的内容外，还分析索引所有指向该网页的链接的URL、AnchorText、甚至链接周围的文字。常用的搜索引擎有Yahoo、Google、新浪、搜狐等。分类指南是将搜索到的网页按照主题内容或学科范围组织成等级结构（主题树），用户按照这个目录逐层深入，直到找到所需信息。通常搜索引擎与分类指南是结合在一起的。

网络学术资源学科导航：是对因特网上的开放信息加以甄别、筛选和科学整理，按学科组织起来，构成的一个完整的学科导航系统，为教学、科研、技术人员提供各类学术信息。

FTP资源：FTP（File Transfer Protocol，文件传送协议），是因特网上最早应用的协议之一，它允许用户登录到远端计算机上，把其中的文件下载到本地的计算机上，或把本地计算机上的文件上传到远端计算机上。所谓FTP资源，指因特网上的开放FTP站点，这些站点允许用户登录上去，下载各类数据、资料、软件等。

其他：如网站、新闻组等。

第二节 医药电子资源的检索

医药电子资源的检索，是指通过一定的检索系统，采用一定的技术手段，根据一定的检索规则，在数据库或其他形式的网络信息资源中找出用户所需的医药学信息资源的过程。简单地说，它是一个信息获取（information access）的过程，是在人、计算机和网络的共同作用下完成的。

一、电子信息资源检索的原理

电子信息资源检索是在手工检索、半机械检索、机械检索、光电检索的基础上演化而来的，其信息检索的本质没有变，但信息的表示方式、存贮结构和匹配方法发生了变化，即要用计算机可以识别的代码来表示信息，用便于计算机快速存取的方式存贮信息，信息的匹配方法也由人工匹配变为机械匹配，匹配方式由隐式变为显式。在这种机械匹配过程中，原先在人工匹配中用于表达概念的语词符号变为没有内涵的字符串。因此，电子资源的检索原理就是将表示用户需求的字符串与计算机内存贮的大量字符串（信息资料的集合）进行比较和逻辑运算的过程。若二者一致或部分一致，并符合给定的逻辑运算条件，即为命中，然后将命中的信息输出给用户。综上所述，电子信息资源检索的原理可归纳为：将检索提问标识与系统中存贮文献的特征标识进行比较，并输出命中文献，检索过程为字符串匹配和逻辑运算的过程。

二、电子信息资源检索系统的构成

电子信息资源检索系统从物理构成来讲,可以分为硬件、软件、数据库三部分。

硬件(hardware):也称为硬件环境,是和计算机检索有关的各种硬件设备的总称,如大型计算机主机(服务器)、存储器(硬盘或光盘)、网络(广域网、局域网或存储区域网等)、输入输出设备(键盘、打印机、鼠标等)、计算机终端或个人计算机(PC机)等。

软件(software):指与计算机检索相关的数据库系统软件及相关应用软件。包括信息采集、存储、信息标引加工、建库、词表管理、用户检索界面、提问处理、网络发布、数据库管理等模块。随着网络和计算机技术的发展,软件的开发平台、程序语言的持续升级,用户功能需求的增加,这一部分的具体结构在不断发生变化。如MEDLINE光盘的检索系统软件WinSPIRS、WebSPIRS等。

数据库(database):指按一定方式、以数字形式存储,可以通过计算机存储、相互关联的数据集合。数据库由文档构成,文档由记录构成,记录由字段构成。文档是记录的有序集合。记录是构成数据库的信息单元,在信息检索数据库中,一条记录代表一篇文献。字段是组成记录的数据项,信息检索数据库中的字段反映一篇文献的具体特征,如标题字段、著者字段、文献来源字段、主题词字段、关键词字段、文摘字段、语种字段等,每个字段都有字段标识符供识别。

三、情报检索语言和自然语言

从电子信息检索的原理可以看出,信息检索的全过程包括存储和检索两个过程。存储过程是指建立检索系统;检索过程即是利用检索系统来查找所需的信息。存储信息时,文献标引人员首先要对文献进行主题分析,即把它所包含的情报内容分析出来,使之形成若干能代表文献主题的概念,并用情报检索语言的语词(标识)把这些概念标示出来,然后纳入检索系统。用户要检索到适当的信息,关键是要使自己使用的提问标识与计算机检索系统内存储使用的标识一致。二者达到统一的关键在于使用相同的检索语言。因此,检索语言在检索过程中的作用非常重要。

情报检索语言(information retrieval language),又称标引语言、索引语言、信息存储与检索语言,是根据情报检索的需要而创建的人工语言,专门用于各种手工的和计算机化的情报检索系统,表达文献主题概念和检索课题概念。它作为提供文献内容检索途径的情报检索系统的一个构成因素,在其中起着语言保证的作用。每种情报检索语言都是由一整套概念标识构成的一个概念标识系统。所谓文献标引,是指用某种概念表示系统的标识来表达文献主题概念的操作过程。

所谓人工语言,是人类根据自定的一套规则编制成的,其规则是预先明确规定的。而自然语言则是人类社会集团在交际中自然演变形成的,其规则反映预先未被明确规定的社会流行用法,是约定俗成的。

(一) 情报检索语言的类型

情报检索语言根据其结构原理,可分为分类检索语言、主题检索语言和代码检索语言。此外,还有一种引证关系追溯法,按其作用来看,也可以说是情报检索语言的一种

类型。

1. 分类检索语言 分类检索语言是用分类号来表达各种概念,将各种概念按学科性质进行分类和系统排列。它又分为体系分类语言和组配分类语言。体系分类语言是运用概念划分的方法,按文献所属学科专业性质的逻辑次序,以分类号为标识,用来存储、检索文献的一种标识系统。组配分类语言一般按学科性质分组,称为“组面”,标引时选择“组面”和有关词汇加以组配。它是一种新型分类方法,其原理是用有关词汇结合文献信息的内容加以组配。

国内外比较重要的分类语言表(分类法)有《国际专利分类表》、《杜威十进分类法》、《中国图书馆分类法》、《中国科学院图书分类法》等。分类表的构成,大致可分为以下几部分:编制说明、基本大类表、简表、详表、辅助表、索引等。

分类检索语言的特点为:

- (1) 分类法符合人们认识事物的规律和处理事物的习惯,因此容易被人们所掌握。
- (2) 由于分类法是按学科、专业集中文献,能系统地揭示文献内容特征,因此,从学科或专业的范围检索文献,能够实现族性检索,获得较高的查全率。
- (3) 便于组织图书资料的排架。
- (4) 查准率不高,主要是受类目数量的限制,不能满足检索专指度较高的文献。
- (5) 由于不同检索系统的分类检索语言使用了不同的分类法,检索标识不同,在检索文献时,要将检索需求的主题概念转化成分类号,容易出错,且不易掌握。
- (6) 分类检索语言是一种先组式语言,不能随时修改补充。因此,新兴学科和边缘学科一旦出现,往往会排列在意想不到的类目下,造成检索上的困难。

2. 主题检索语言 主题检索语言用语词来表达各种概念,将各种概念不管其相互关系完全按字顺排列。主题语言有两大特点:①直接用能表达和描述文献内容特征的名词性属性作为标识来揭示文献的内容特征;②把这些标识按字顺排列成主题词表,以此作为标引和检索文献信息的工具。主题语言是一种描述性语言,即用自然语言中的名词、名词性的词或词组描述文献所论述或研究的事物概念(主题),文献信息的主题则是文献研究、讨论、阐述的具体对象或问题,两者有本质区别。主题词是主题语言的核心,是用以描述文献主题概念的名词性术语,这些名词性术语取自自然语言,有的经过规范化处理、有的本身就是自然语言中的一部分。标题词语言、单元词语言、叙词语言都属于主题语言。

用主题词作为文献标识具有以下优点:

- (1) 直观:主题词来源于自然语言,标识比较直观,符合人们的辨识习惯。主题词在词表中按词的字顺排列,序列明确,易于利用。只要会使用字典、词典,一般都能很顺畅地使用这种语言。
- (2) 专指性强:用做主题词的语词标识经过了全面严格的规范化处理,一个标识与一个概念严格对应,标识和所表达概念的唯一性(即单义性)使语词对概念的描述具有专指性。
- (3) 灵活:通过词与词之间的概念组配来揭示文献中形形色色的主题,是主题语言的主要特征,尤其是后组式的组配原则,便于人们按照检索需要,自由组配检索概念,具有很大灵活性。

(4) 网罗度高: 一个主题词表达一个泛指的事物概念, 若干个主题词用一定的逻辑方法进行的组配, 可形成高度专指的概念特征。用于标引文献时, 概念网罗度高。把主题词按照一种利于检索的方式 (通常是字顺) 编排成册, 就形成主题词表, 它是主题标引的主要工具。主题词表实际上就是主题词加有某种参照的字顺体系表, 它是将自然语言转换为文献信息检索语言的术语控制工具。由于主题词表列举的概念标识数量较多, 多数标识的指代范围较窄, 因此, 使用主题语言检索文献具有直指性强、专指度高的特点。检索者不必从知识体系的角度去判断所需文献属于什么学科, 只要根据课题研究对象, 直接用能描述文献内容主题词去查找。而且同一篇文献可以用多个主题词来标引, 因此扩大了检索的途径, 在这一点上, 主题语言比分类语言具有更大的优越性。

3. 代码语言 代码语言是指针对事物的某方面特征, 用某种代码系统来表示和排列事物概念, 从而提供检索途径的检索语言。例如, 根据化合物的分子式这种代码语言可以构成分子式索引系统, 允许用户从分子式出发, 检索相应的化合物及其相关的信息。

(二) 自然语言

自然语言是对同义词、多义词、近义词等不加处理, 检索用词是从信息内容本身抽取的, 主要依赖于计算机自动抽词技术完成, 辅以人工自由标引 (不依据词表)。自然语言在情报检索中的应用是以计算机检索为前提的, 不使用计算机, 自然语言检索就难以实现。自然语言在情报检索中应用的方式很多, 可大体分为不标引方式、自动抽词方式、关键词索引方式、人工自由标引方式、自然语言与人工语言并用方式和自然语言接口方式六类。

其中关键词索引法在国外曾得到广泛的应用。关键词索引法简称关键词法, 就是将文献中原有的、能描述其主题概念的关键词抽出, 不加规范化处理, 按字顺排列, 以提供检索途径的语言。所谓关键词, 是指那些出现在文献的标题 (篇名、章节名) 以及摘要、正文中, 对表达文献主题内容来说是最重要的、关键性的、可作为检索入口的语词。关键词索引是计算机最早编制的检索工具之一。关键词语言适用于计算机自动化编制关键词索引, 故缩短了检索工具出版的时间, 由于关键词使用的是自然语言, 故容易掌握, 使用方便。关键词语言的缺点是查准率和查全率较低, 这是因为关键词是未经规范化处理的自然语言, 其同义词、近义词未加规范统一, 就会造成标引与检索之间的误差, 导致文献的漏检。另外, 自然语言中多种形式的相关关系在关键词中得不到显示, 这也给查准查全文献带来困难。

目前情报检索语言与自然语言在同一个检索系统中并用, 情报检索语言增加自然语言成分, 自然语言适当引进情报检索语言的原理与方法。自然语言或情报检索语言的将来将是自然语言的情报检索语言化或情报检索语言的自然语言化, 是两者互相结合到完全融合的过程, 这是自然语言和情报检索语言发展的大趋势。

第三节 医药电子资源的检索途径、检索技术与检索步骤

一、检索途径

如前所述, 数据库由若干条记录集合而成, 而一条记录又包含若干字段。一篇文献

用一条记录来揭示,一条记录又包含分类、题名、著者、出处、年份、(题中)人名、主题等字段,其中大部分字段都是可检索字段,这些可检索字段,我们就称之为检索途径。

一般来说,每个数据库都有多个检索途径,常用的有以下几种:

(一) 主题词检索

主题词检索是根据标引人员按照规范词表标引出的主题词进行检索的途径,其优点一是能满足特性检索要求,专指性强;二是能适应新兴学科及多学科文献检索,只要根据新学科的出现、发展及多学科的需要,随时增加主题词,就能快速检出所需文献。其缺点是主题词选择必须准确,否则无法进行查找;另外由于主题词规范性的要求,输入的主题词必须完全正确,因此每次使用时需要查找主题词表,较费时间。日前国内外采用主题词检索的数据库,其使用的主题词表均具有主题词检索和分类检索的特点,因而也可适应检索范围大、内容广的课题。

主题词由主题词表来进行规范,现在应用广泛的一个医药学方面的主题词表是美国国立医学图书馆(National Library of Medicine,简称 NLM)编辑的《医学主题词表》(Medical Subject Headings,简称 MeSH),例如,有关“艾滋病”的描述,著者可以用艾滋病、AIDS、获得性免疫缺陷综合征等词表达,但是 MeSH 规定,只有获得性免疫缺陷综合征才是主题词,因此用主题词途径检索的时候一定要用规范的主题词才能检索到所需要的文献信息。

很多数据库都有主题词检索途径,如中国生物医学文献数据库、MEDLINE、PubMed 等等。

(二) 分类检索

分类法是利用学科、专业、概念之间的逻辑关系建立的一种等级体系。在分类法中,用数字或者数字加字母构成的分类号代表一个一个的概念,这些概念能直接反映上位类与下位类之间的从属关系,也能反映同位类之间的并列关系。分类号检索根据按一定规则编排的分类表进行检索。其优点是能满足族性检索的要求,便于查全某一学科或某一专业的文献;易于扩大或缩小检索范围,扩大时可取上位类,缩小时可取下位类。

我国图书和资料按照《中国图书馆分类法》进行分类。例如,R 医药卫生、R28 中药学、R5 内科学、R9 药学。在中国期刊全文数据库、中国生物医学文献数据库等数据库都有分类检索途径。

(三) 关键词检索

关键词检索的最大优点是词语不必规范化,用户可根据自己的需要,选择熟悉的词语进行检索,不用特意记忆或事先查找词汇,比较方便。其缺点是容易漏检,因而使用这种途径进行检索时,必须同时考虑多个同义词、近义词,以减少漏检。在不同的数据库中关键词检索字段的名称可能不同,如 EBSCO 数据库使用的是:tx All Text,指在所有可以检索的字段里进行关键词检索;OVID 检索平台使用的是:kw Author Keywords 和 tw Text Word 两个字段。

值得注意的是,某些数据库中的关键词检索并不是检索记录中的任意文本词,而是只检索原文著者给出的数个关键词,例如中国期刊全文数据库。

(四) 著者检索

著者检索是根据文献信息的外部特征,利用著者名进行检索的途径。国外比较重视著者途径,许多数据库都把著者检索作为最基本的检索途径之一,它是按著者的姓名字顺查找文献的一种方法。以著者为线索可以系统、连续地掌握他们的研究水平和研究方向,同一著者的文章往往具有一定的逻辑联系,著者途径具有一定的族性检索功能。已知课题相关著者姓名,便可以依著者途径迅速准确地查到特定的信息,因此亦具有特性检索的功能。

在中文数据库中直接用著者的中文姓名检索即可。在英文数据库中查找国外著者的时候,注意将姓名倒装,名字用缩写。例如,查找 John Smith 的文献,一定要用 Smith, J. 来查找。查找国内的著者用著者的汉语拼音,也是姓在前,名在后,名字用缩写。但是利用国外数据库查找国内著者时,要注意著者姓名的著录有时不规范,要参照机构和文章内容加以鉴别。

(五) 引文检索

在学术论文、综述、专著文献后面,通常列有著者借鉴、利用过的文献信息,称之为参考文献或引文。这里,列有参考文献的文章叫做“引用文献”,列出的参考文献称为“被引用文献”,引文检索是以被引用文献为检索起点来查找引用文献的过程。如果某一作者、某一机构,或某一刊物想查找自己所著、所出的文献被引用的情况,可以运用引文检索。

目前引文检索报告已经作为作者评定职称、申报成果的重要参考证明,要求提供引文检索服务的人也越来越多。引文从一个侧面反映科技成果被他人的认可度、利用度和学术价值。通过引文检索途径,可对文献中引用的参考文献的相关内容进行检索;通过对某个作者、某个出版物、某篇文章进行统计分析,可对其做出评价。世界最著名的引文数据库为 Web of Science,它包括了 SCI (Science Citation Index),国内的医学引文数据库有中国生物医学期刊引文数据库 (CMCI)。

(六) 机构检索

机构检索是以机构名称为检索词,来查找该机构学者发表的文献。不少数据库把机构名并入著者地址字段,例如美国的 MEDLINE。用机构检索途径是统计某一机构或者单位发表文献数量最好的途径。需注意的是有些数据库中机构名或者单位名是用缩写形式著录的,因此检索时也要使用机构或单位名称的缩写形式。

(七) 刊名检索

刊名检索是提供检索指定刊物上发表文献的途径。有的数据库提供刊名浏览,简单点击刊名链接即可,有的则要输入期刊的刊名。西文期刊名的输入有刊名全称和缩写两种情况,两者不能混淆。利用 MEDLINE 的 SERFILE 文档可以通过刊名缩写查到刊名全称。

(八) 序号检索

有些文献有特定的序号,如专利号、报告号、合同号、标准号、国际标准书号和刊号等。文献序号对于识别一定的文献,具有明确、简短、唯一性特点,依此编成的各种序号索引可以提供按序号自身顺序检索文献信息的途径。

电子信息资源检索途径的另一种划分方法是把检索途径分为三种:一是简单检索