

高等医药院校教材
供基础、临床、预防、口腔医学类专业用

医学数字资源的 检索与利用

第3版

主 编 | 马 路

副主编 | 黄 芳 田新玉



人民卫生出版社

高等医药院校教材
供基础、临床、预防、口腔医学类专业用

医学数字资源的 检索与利用

第3版

主 编 马 路

副主编 黄 芳 田新玉

编 者 (以姓氏笔画为序)

马 路	首都医科大学	陈文凯	广东医学院
王 凌	首都医科大学	周 纯	中山大学
田 瑞	首都医科大学	徐 伟	哈尔滨医科大学
田新玉	北京大学第三医院	黄 芳	首都医科大学
刘玉婷	首都医科大学	龚佳剑	首都医科大学
李 宁	首都医科大学	程艾军	首都医科大学
李永杰	首都医科大学	蔡庆芳	广州医科大学

秘 书 丁思瑛 首都医科大学

人民卫生出版社

图书在版编目(CIP)数据

医学数字资源的检索与利用 / 马路主编. —3 版. —北京:
人民卫生出版社, 2013.8

ISBN 978-7-117-17687-3

I. ①医… II. ①马… III. ①医学—情报检索—医学
院校—教材 IV. ①G252.7

中国版本图书馆 CIP 数据核字(2013)第 147252 号

人卫社官网	www.pmph.com	出版物查询, 在线购书
人卫医学网	www.ipmph.com	医学考试辅导, 医学数 据库服务, 医学教育资 源, 大众健康资讯

版权所有, 侵权必究!

医学数字资源的检索与利用

第 3 版

主 编: 马 路

出版发行: 人民卫生出版社(中继线 010-59780011)

地 址: 北京市朝阳区潘家园南里 19 号

邮 编: 100021

E - mail: pmph@pmph.com

购书热线: 010-59787592 010-59787584 010-65264830

印 刷: 北京市文林印务有限公司

经 销: 新华书店

开 本: 787×1092 1/16 印张: 13

字 数: 324 千字

版 次: 2006 年 2 月第 1 版 2013 年 8 月第 3 版

2013 年 8 月第 3 版第 1 次印刷(总第 5 次印刷)

标准书号: ISBN 978-7-117-17687-3/R · 17688

定 价: 28.00 元

打击盗版举报电话: 010-59787491 E-mail: WQ@pmph.com

(凡属印装质量问题请与本社市场营销中心联系退换)

前言

《医药电子资源的检索与利用》2006 年出版了第 1 版, 2009 年出版了第 2 版, 自出版以来受到了广大教师与学生的欢迎与支持, 我们深表感谢。医学信息检索是一门创新课程, 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。随着社会的进步与信息化时代的到来, 医学信息检索的发展速度更是十分惊人, 为此我们在前两版的基础上对本书进行了修订。为了增强实用性, 更有利于学生学习和掌握, 本书增加了新的内容, 淘汰了一些实用性相对较差的部分, 主要变化如下:

1. 全书将电子资源改称为数字资源。

2. 根据目前的出版情况和用户使用情况对数据库内容进行了删改, 例如循证医学资源及证据检索中新增了 BEST PRACTICE 和 EVIDENCE-BASED MEDICINE REVIEWS 数据库, 在专类文献信息检索中删除了科技报告及标准文献信息检索, 增加了学位论文、会议文献和专利文献的概念及分类, 增加了 Springer Protocol、MICROMEDEX 事实数据库、ClinicalKey、Primal Pictures 3D 解剖学数据库和好医生医学点播课堂等。

3. 为了加强实用性, 在信息利用与分析中新增了信息分析方法一节。

4. 删除了 Internet 核酸序列检索。

在本书的编写过程中, 我们参考了大量有关的专著和文献, 并得到了首都医科大学文献检索教研室的大力支持, 在此谨向这些老师和作者表示衷心的感谢!

由于医学信息检索的理论和实践研究的不断深入和飞速发展, 许多问题有待进一步探讨, 尽管本书的编委大多具有高级职称或硕士学位, 并都具有丰富的科研和教学经验, 但平均年龄仍旧比较年轻, 编写水平有限, 加之时间仓促, 难免会有疏漏和错误之处, 因此恳请读者批评指正。

马 路

2013 年 4 月 30 日

目 录

第一章 医学数字资源的检索与利用概论	1
引言	1
第一节 医学数字资源的概念与类型	2
一、数字资源的概念与特性	2
二、数字资源的类型	3
第二节 医学数字资源的检索	5
一、数字资源检索的原理	5
二、数字资源检索系统的构成	5
三、检索语言	6
第三节 医学数字资源的检索途径、检索技术与检索步骤	8
一、检索途径	8
二、检索技术	10
三、检索步骤	11
第四节 如何合理使用网络	13
一、大学生在利用信息和使用网络过程中经常出现的问题	13
二、遵守网络道德,合理使用网络	14
思考题	14
第二章 参考型数据库	15
第一节 中国生物医学文献数据库	15
一、数据库简介	15
二、检索途径和方法	16
三、检索结果的处理	23
四、原文索取	24
五、检索实例	25
第二节 中文生物医学期刊文献数据库	25
一、数据库概况	25
二、检索途径和方法	25
三、检索结果的处理	30
四、检索实例	31
五、CBM 和 CMCC 的比较	31

第三节 PubMed 数据库	32
一、数据库概况	32
二、检索途径	36
三、检索结果处理	42
四、检索实例	43
第四节 EMBASE.com 数据库	44
一、数据库概况	44
二、检索途径和方法	45
三、检索结果处理	49
四、检索实例	50
第五节 BIOSIS Previews 数据库	50
一、数据库概况	50
二、检索途径	53
三、检索结果处理	57
四、检索实例	59
第六节 其他国外参考数据库	59
一、化学文摘数据库(CA)	59
二、国际药学文摘(IPA)	60
三、工程索引(EI)	60
四、卫生保健及护理学数据库(CINAHL)	60
五、肿瘤文献数据库(CancerLit)	61
思考题	61
第三章 全文数据库	62
第一节 中国知网	62
一、数据库概况	62
二、检索途径和方法	62
三、检索结果的处理	66
四、检索实例	67
第二节 万方数据资源系统	67
一、数据库概况	67
二、检索途径和方法	68
三、检索结果的处理	68
四、检索实例	69
第三节 维普信息资源系统	69
一、数据库概况	69
二、检索途径和方法	70
三、检索结果的处理	72
四、检索实例	73

第四节 ScienceDirect	74
一、数据库概况	74
二、检索途径和方法	74
三、检索结果的处理	78
四、个性化功能	79
第五节 EBSCOhost	81
一、数据库概况	81
二、检索途径和方法	83
三、检索结果的处理	87
四、个性化功能	88
第六节 Ovid 全文电子期刊	88
一、数据库概况	88
二、检索途径和方法	89
三、检索结果的处理	91
四、个性化功能	92
第七节 PHMC 健康与医学大全	93
一、数据库概况	93
二、检索途径和方法	93
三、检索结果的处理	95
四、个性化功能	96
第八节 其他国外全文数据库	97
一、Wiley Online Library	97
二、SpringerLink 全文数据库	98
三、Karger 医学电子期刊	98
四、BMJ 期刊专辑	98
五、OUP 电子期刊	98
第九节 电子图书	99
一、读秀学术搜索	99
二、方正 Apabi 数字图书馆	100
三、书生数字图书馆	100
四、Thieme 医学电子图书	101
五、Ebrary 电子图书	102
六、Karger 电子图书	102
七、World eBook Library	103
第十节 开放获取期刊	103
一、开放获取及开放获取期刊	103
二、国外开放获取期刊资源	104
三、国内开放获取期刊资源	107
思考题	108

第四章 引文索引	109
第一节 Web of Science	110
一、概况	110
二、数据库结构与检索规则	111
三、检索途径与方法	113
四、保存检索历史和创建定题跟踪服务	118
五、检索结果的处理	118
六、检索综合举例	121
七、Web of Science 应用	121
第二节 中文引文检索工具	125
一、中国生物医学期刊引文数据库	125
二、中国科学引文数据库	126
第三节 其他检索工具的引文检索功能	126
一、中文数据库的引文检索	126
二、英文数据库的引文检索	127
思考题	128
第五章 专类文献信息检索	129
第一节 学位论文检索	129
一、学位论文概述	129
二、国内学位论文数据库	130
三、国外学位论文数据库	132
第二节 会议信息检索	135
一、会议文献概述	135
二、会前信息检索	136
三、会后信息检索	138
第三节 专利文献检索	140
一、专利基础知识	140
二、专利文献基础知识	143
三、国际专利分类表	147
四、中国专利文献检索	148
五、国外专利数据库	152
六、综合性专利检索系统	158
七、其他专利文献检索平台	159
第四节 其他专类数据库	159
一、Springer Protocol	159
二、MICROMEDEX 事实数据库	160
三、ClinicalKey	161
四、Primal Pictures 3D 解剖学数据库	161

五、好医生医学点播课堂·····	161
思考题·····	162
第六章 循证医学资源及证据检索·····	163
第一节 循证医学简介·····	163
第二节 循证医学证据的检索·····	164
一、一级来源证据·····	164
二、二级来源证据·····	164
第三节 Cochrane 协作网和 Cochrane Library ·····	167
一、Cochrane 协作网的历史 ·····	167
二、Cochrane 协作网的发展及现状 ·····	167
三、The Cochrane Library ·····	168
第四节 Best Practice ·····	171
一、概况·····	171
二、使用方法·····	172
第五节 Evidence-Based Medicine Reviews ·····	175
一、概况·····	175
二、使用方法·····	175
思考题·····	177
第七章 信息利用与分析·····	178
第一节 检索策略·····	178
一、检索步骤·····	178
二、检索策略的调整·····	179
三、实例·····	181
四、跨库检索平台·····	182
五、全文传递与馆际互借·····	184
第二节 个人文献管理软件·····	189
一、NoteExpress ·····	189
二、Endnote ·····	192
第三节 信息分析方法·····	193
思考题·····	198
参考文献·····	199

第一章

医学数字资源的检索与利用概论

引言

近年来,随着我国高校素质教育工作的不断深入,信息素质(information literacy)教育日益受到重视。信息素质这一概念,最早是由美国波尔(Paul Zurkowski)于1974年提出,并定义为“利用大量的信息工具及主要信息源使问题得到解答的技术和技能”。1983年,美国信息学家霍顿(Horton)认为教育部门应开展信息素质教育,以提高人们对联机数据库、通讯服务、电子邮件、数据分析以及图书馆网络的使用能力。

信息素质既是一种能力素质,更是一种基础素质。信息素质有其自身的内容结构,它包括信息意识素质、信息能力素质和信息道德素质。信息意识是人们对信息的自觉反应,是感觉、思维等各种心理过程的总和,是人的社会意识的一部分。信息能力,即人们搜集、贮存、处理、探索、利用和传递信息的能力。信息道德是指整个信息活动中的道德,它是信息生产者(拥有者)、信息服务者及信息使用者必须遵循的行为准则和规范。

美国1989年成立了由美国图书馆协会领导的信息素质委员会,并提出“信息素质是信息社会人的生存能力之一”的重要观点。1998年,美国医学院联合会发布的“医学院目标计划”(MSOP),详细阐述了对医学毕业生的信息素质要求。医师在职业生涯中将扮演“终生学习者、临床医师、教育者/交流者、研究者以及管理者”的角色,医学信息对于这些角色至关重要。医学信息素质教育的目标是使医学毕业生“有能力通过数据库或其他资源检索、管理、运用生物医学信息,解决医疗问题,作出正确决策”。

2001年美国大学和研究图书馆协会(ACRL)制定的“高等教育信息素质能力标准”提出的5个标准,既是大学信息素质教育的目的,也是考核大学生信息素质能力的指标。

标准一:学生能明确信息需求的内容与范围。具体评价标准包括:①明确了解信息需求;②识别各种类型和格式的信息源;③能评价获得信息的成本与效益;④能重新评价所需信息与范围。

标准二:学生能高效获取所需信息。具体评价标准包括:①能选择适当的调查方法或信息检索系统获取所需信息;②能构建与实施有效的检索策略;③能利用联机检索终端或采用不同的方法检索所需信息;④必要时能调整检索策略;⑤能提取、记录和管理信息及信息源。

标准三:学生能客观、审慎地评价信息与信息源,利用知识基础和价值系统选择适当的

信息。具体评价标准包括：①能从获取的信息中提炼信息主题；②能采用有关标准评价信息及信息源；③能综合信息要点形成新的概念；④能通过新旧知识的比较而确定信息的增加值；⑤能确定新的知识对个人价值体系的影响，并使其融合于个人的价值体系中；⑥能通过与他人的交流了解自己是否能表达所获取的信息；⑦能决定是否修订初始的查询方式。

标准四：学生作为个人或作为群体一员能有效利用信息实现特定的目标。具体评价标准包括：①能够运用新的和以前的信息开发新的产品或项目；②能调整开发产品或项目过程；③能有效地将开发的产品或项目情况与他人沟通。

标准五：学生能够理解有关信息使用的经济、法律以及社会因素，获得和使用信息时能自觉遵守道德规范和有关法律。具体评价标准包括：①了解信息与信息技术使用的相关道德、法律、经济和社会问题；②在存取、使用信息资源时能够遵守法律法规和信息资源提供的规定以及约定俗成的一些规则；③对引用的成果表示致谢。

本书参考上述的标准二，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 医学数字资源的概念与类型

一、数字资源的概念与特性

数字资源(digital resources)，也叫电子资源，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。数字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送出去，最终显示在用户的计算机终端上。医学数字资源是指包含医药学相关知识的数字资源。

随着互联网的发展，利用网络传递的数字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数字资源(network digital resources)。除此之外，目前仍然存在着仅在本地上使用、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，也属于数字资源范畴。

数字资源的特性为：

1. 易获取性 数字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，互联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数字资源。数字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，数字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。而数字资源则不同，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，体现了数字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 多媒体技术在数字资源中的利用,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不充分和检索手段烦琐的限制,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、数字资源的类型

医学数字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。

(一) 按照载体形态划分

数字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数字资源;二是在网络上出版的网上数字资源。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,但随着互联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,网上资源已成为图书馆数字资源的主体,光盘只是作为数据备份的辅助工具。

(二) 按照性质和功能划分

1. 一次文献(primary literature) 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。

2. 二次文献(secondary literature) 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。

3. 三次文献(tertiary literature) 研究人员围绕某一专题,借助于二次文献,在充分研究和利用大量一次文献的基础上,经过阅读、分析、归纳、概括,撰写而成的新文献就是三次文献。其形式有综述、述评、进展、现状、发展趋势等期刊文献和百科全书、年鉴、手册等参考工具书。

(三) 按照发布形式划分

1. 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,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。

2. 全文数据库(full-text database) 即收录有原始文献全文的数据库,包括经典著作、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。最主要的类型有电子期刊、电子图书和电子报纸。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,只是两层用特定化学物质作为透明涂层的薄膜,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。薄膜带负电的部分因吸附带正电的黑粒而呈现黑色;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方式发送到你的阅读器中。

3. 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。

4. 搜索引擎/分类指南(search engines/classification guide) 搜索引擎通常是指收集了互联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成为搜索结果。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 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。

5. 网络学术资源学科导航(Internet academic resources navigation) 是对互联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。

6. FTP 资源 FTP(file transfer protocol, 文件传送协议)是互联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,指互联网上的开放 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。

7. 其他 如网站、新闻组等。

第二节 医学数字资源的检索

医学数字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、数字资源检索的原理

数字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,数字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。若两者一致或部分一致,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,数字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、数字资源检索系统的构成

数字资源检索系统从物理构成来讲,可以分为硬件、软件、数据库3部分。

(一) 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。

(二) 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、信息索引加工、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。

(三) 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、检索语言

从计算机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,即把所包含的信息内容分析出来,使之形成若干能代表文献主题的概念,并用信息检索语言的语词(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。两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的语言,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语言,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) 检索语言的类型

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的主题词去查找,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,检索用词从信息内容本身抽取,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6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 医学数字资源的检索途径、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:①能满足特性检索要求,专指性强;②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。目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获得性免疫缺陷综合征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、PubMed 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各种的概念,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。