

医学信息检索

YIXUEXINXIJIANSUO

主编 姜振儒 翟俊霞



内蒙古科学技术出版社

责任编辑: 刘 冲
封面设计: 姜 一

医学信息检索

YIXUEXINXIJIANSUO

ISBN 978-7-5380-1781-6



9 787538 017816 >

定价: 23.80 元

医学信息检索

主 编 姜振儒 翟俊霞

内蒙古科学技术出版社

图书在版编目(CIP)数据

医学信息检索 / 姜振儒, 翟俊霞主编. — 赤峰: 内蒙古科学技术出版社, 2009. 1

ISBN 978 - 7 - 5380 - 1781 - 6

I. 医… II. ①姜…②翟… III. 医药学—情报检索
IV. G252. 7

中国版本图书馆 CIP 数据核字(2009)第 009332 号

出版发行: 内蒙古科学技术出版社

地 址: 赤峰市红山区哈达街南一段 4 号

电 话: (0476)8224848 8231924

邮 编: 024000

出 版 人: 额敦桑布

责任编辑: 刘 冲

封面设计: 姜 一

印 刷: 赤峰地质宏达印刷有限责任公司

字 数: 240 千

开 本: 787 × 1092 1/16

印 张: 10.25

版 次: 2009 年 1 月第 1 版

印 次: 2009 年 1 月第 1 次印刷

定 价: 23.80 元

编委会

主 编：姜振儒 翟俊霞

副主编：万丽娟 刘玉华 刘丹丹

编 者（按姓氏笔画排列）

万丽娟 王 韪 华 刘丹丹

刘玉华 李 达 张松岩

姜振儒 胡艳君 阎 敏

翟俊霞

内 容 简 介

本书主要介绍了网络医学信息检索系统及数据库,药学信息检索系统及数据库,循证医学信息检索系统及数据库,以及医学信息查新等。本书既可作为高等医学院校研究生、本科生医学信息检索课的教材,又可作为医学教学、科研、医务人员继续教育的材料。

前言

进入 21 世纪,随着网络技术和数字出版技术的迅猛发展,医学信息资源的检索与利用正面临一些新问题:①网络信息资源大量增加,数据质量参差不齐,检索平台五花八门,使用户对信息资源的甄别与取舍更加困难;②用户检索需求日益深化,不仅需要整本、整套的信息资源,也需要具体的篇章、图片、镜头和片断;③用户对数字资源依赖程度越来越强,通过网络查找信息已成为人们获取信息的首选方式;④网络信息搜索引擎不断更新、网络信息检索环境更加友好。针对这些问题,医学信息检索课必须紧跟时代的发展,随时更新教材内容,满足大学生对网络信息检索的需求。

因此,在总结近几年来教学实践的基础上,结合国内已有的比较好的教材,大量增加了网络医学信息检索的内容,并根据各种数据库的新变化,对原教材相应内容进行了修改。新教材具有以下几个特点:①内容新,许多数据资料截止 2008 年 11 月;②实用性强,本次新收入的检索工具和数据库多数是平时经常使用的或是能在网上免费查询的;③编排合理,将传统的手工检索工具和与其配套的电子版数据库放在一起编排,如美国《IM》与 MEDLINE,荷兰《EM》与 EMBASE 等,使学生的学习时能由浅入深,相互对照,加深理解;④形象直观,对重点内容采用原件或图片示例,图文并茂,便于学生自学。

本书共分八章,内容包括医学信息检索的基本理论和基本知识、中文医学信息检索工具及数据库、外文医学信息检索工具及数据库、网络医学信息检索系统及数据库、药学信息检索工具及数据库、循证医学检索系统及数据库、特种医学文献检索和医学查新咨询等。

本书编者来自承德医学院、河北省医学情报研究所和牡丹江医学院,是长期从事医学信息检索教学和医学信息查新服务的专业人员。第一章由姜振儒编写,第二章由刘丹丹编写,第三章由万丽娟、胡艳君编写,第四章由翟俊霞、张松岩编写,第五章由李达编写,第六章由王韞华编写,第七章由阎敏编写,第八章由刘玉华、翟俊霞编写。在编写过程中参考了最近两年出版的同类教材及相关论著,从中受益匪浅。但由于水平所限,书中难免存在纰漏,敬请同行专家、老师及学生们批评指正。

编者

2008 年 11 月

目 录

第一章 绪论	1
第一节 信息、情报、知识、文献	1
第二节 信息资源及其类型	3
第三节 信息检索	5
第四节 计算机检索概述	10
第二章 国内医学信息检索工具及数据库	19
第一节 中国生物医学文献数据库(CBMdisc)	19
第二节 中国期刊全文数据库	23
第三节 万方数据医药信息资源系统	28
第四节 维普资讯网	31
第五节 其他国内医学信息检索工具	33
第三章 国外医学信息检索工具及数据库	39
第一节 美国《医学索引》	39
第二节 MEDLINE 数据库	48
第三节 荷兰《医学文摘》	53
第四节 EMBASE.COM	56
第五节 美国《生物学文摘》	62
第六节 BIOSIS Preview 数据库	65
第七节 美国《化学文摘》	68
第八节 CA on CD	71
第九节 美国《科学引文索引》	73
第十节 Web of Science	78
第四章 网络医学信息检索系统及数据库	83
第一节 PubMed 数据库	83
第二节 GoPubMed	93
第三节 Ovid 全文期刊数据库	98

第四节 NLM Gateway 检索平台	100
第五节 MEDLINEplus	105
第五章 药学信息检索工具及数据库	109
第一节 常用药学文献检索工具及数据库	109
第二节 Internet 上的主要药学专业网站	112
第六章 循证医学检索系统及数据库	121
第一节 循证医学概述	121
第二节 循证医学信息检索	125
第七章 特种文献检索工具及数据库	130
第一节 专利文献检索	130
第二节 会议文献检索	139
第三节 学位论文检索	142
第八章 医学查新咨询	145
第一节 查新咨询工作概述	145
第二节 查新咨询的作用	147
第三节 查新咨询的类型	148
第四节 查新咨询的程序	149
第五节 查新咨询工作的质量控制	152
参考文献	154

第一章 绪 论

第一节 信息、情报、知识、文献

一、信息(Information)

信息是事物存在和运动状态及其特征的反映,即客观事物所发出的信号和消息。信息普遍存在于自然界、人类社会以及人们的思维活动中。事物在不断变化,信息也在不断产生。不同的事物,不同的运动状态和特征发出不同的信息,例如:物质发出的声、光、形、色、味;人体发生的语言、姿态、行为、表情;细胞染色体的脱氧核糖核酸所含的遗传密码,以及人体的脉搏、体温、呼吸甚至疾病症状的发生和变化等,凡是能被人感知的都是信息。

信息具有客观性、可识别性、时效性、共享性和依附性等基本属性。

信息的客观性,是指物质的客观存在决定了物质及其运动的某种特征,那么体现这种特征的信息自然也是客观存在的,不以人的意志为转移的,所以信息是客观的。

信息的可识别性,是指人类可以通过自己的感觉器官和视觉器官直接认识和识别某些信息,也可以借助各种工具间接识别信息。由于人类科学技术水平的限制,自然界还有大量的信息至今尚未被人们所识别。随着人类科学技术的发展,人们可识别的信息领域会越来越宽广。

信息的时效性,人们获取信息的目的在于利用,信息的效用与利用时间密切相关,使用适时才有最好的价值,使用迟缓就会降低效用甚至对工作造成损害。时效性要求在搜集、加工、提供、利用信息时都必须注意及时、快捷。时效性决定着信息的经济价值、社会价值和学术价值的大小。

信息的共享性,是指信息能为一切信息接受体所共同享用,可以多次重复利用。信息不像一般物质资源,它不属于特定的占有对象,可以为众多的人们共同享用。实物转赠之后,就不再属于原物主,而信息传递给他人之后,自己仍可以利用。

信息的依附性,信息的存储、传递和交流都必须依附于一定的物质载体,而同一信息可以借助不同的载体存在,可以用不同的符号表示,其性质和功能不会发生改变。

二、知识(Knowledge)

知识是人类在认识世界和改造世界的社会实践中所获得的正确认识和经验的总和,是人的主观世界对客观世界的如实反映和概括。知识来源于信息,人类在认识世界和改造世界的过程中,通过接受客观事物的信息,经过大脑的思维加工把感性认识和经验总结成为知识。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,这个例子前者显然是获得的信息,后者则是这类信息组成的神经反射知识。

知识按其内容可分为自然科学知识、社会科学知识和哲学知识。医学知识属于自然科学知识范畴。

三、情报(Intelligence)

情报是关于某种情况的消息和报告,最早指军事上“战时敌方情况之报告”。随着社会的进步和发展,人们对情报的需求愈来愈多,情报的内容也愈来愈广泛,情报交流方式也趋于多样化,情报已经渗透到各行业、领域,成为经济、政治、军事、文化和科学技术向前发展必不可少的条件。关于情报的定义至今尚无定论,比较常见的说法有:“情报就是为了解决一个特定问题所需要的知识,是激活了的知识,以及这种知识的及时性和针对性”;“情报是知识经传递并起作用的部分”。由上述可见,只有被激活、被传递、被利用的知识才是情报,而沉睡的知识不是情报。

情报具有知识性、传递性、效用性三个基本属性。

知识性,即情报的本质是知识,知识是情报的实体。

传递性,知识在交流和传递的过程中,经过被激活、被利用后才能成为情报。

效用性,即情报的应用都应产生一定的效益,是情报使用价值的体现。

四、文献(Literature, Document)

凡是用文字、图形、符号、声频、视频等手段记录下来的人类知识都可以称之为文献,换言之,文献是记录有知识的一切载体。构成文献的三个最基本要素:一是有知识性,文献的实质内容是知识;二是有物质载体,文献只有依附在物质载体(如纸、感光材料或磁性材料等)上面,才能够长期保存和传播;三是有记录手段,文献存储是通过文字、图形、符号、声频、视频等记录手段实现的。

五、信息、知识、情报与文献之间的关系

从信息、知识、情报与文献的定义可以看出,它们之间有着密切的关系。信息是事物存在和运动状态的反映;知识是人类在实践中进行系统化、理论化的信息;情报是被激活了的有用的知识;文献是记录了知识的一切载体。有人把信息比喻为大海,知识为大海中的一片水域,情报是这片水域中的一个小岛,通过这种形象的比喻我们能更好地理解信息包含着知识,知识包含着文献和情报,而知识、情报、信息的主要部分存在于文献之中。

第二节 信息资源及其类型

一、信息资源(Information Resource)

“信息资源”一词源于美国,是随着信息资源管理理论的兴起而产生的术语。国内外关于“信息资源”一词的概念众说纷纭。狭义的理解是指信息内容的本身,广义的理解是指除信息内容外,还包括信息设备、信息人员、信息系统和信息网络等。就信息资源的狭义概念而言,它是反映客观事物的各种信息的总称,也是各种信息的集合体。正如黑龙江人民出版社1992年出版的《新语词大词典》的解释:“信息资源是指可以通过不同方式和渠道获取的,存在于生活、环境中的一切信息。信息资源广泛存在于经济、社会各个领域和部门,是各种事物形态、内在规律和其他事物联系等各种条件、关系的反映”。

当今世界已步入信息化时代。信息化社会的最重要特征是“信息资源”和自然资源一样,是社会发展不可缺少的资源,是一种宝贵的、富有价值的资源。随着社会的不断发展,信息资源对国家和民族的发展,对人们工作、学习和生活都至关重要,成为国民经济和社会发展的主要资源。它的开发和利用是整个信息化体系的核心内容。

二、信息资源的类型

(一)按信息的情报含量与加工程度分

1. 一次信息资源

一次信息资源是以作者本人的工作经历和科研成果为依据而创作的具有前所未有的发明创造或一些新见解、新观点、新数据、新方法的总结。一次信息资源是科学技术发展的标志,其特点是具有创造性、先进性和知识性,是科技人员参考的基本资料。一次信息资源包括

期刊论文、会议资料、学位论文、科技报告、专利文献等。

2. 二次信息资源

二次信息资源是对一次信息资源进行整理、加工而成的检索工具。即把大量的、分散的、无序的一次信息资源收集起来,按照一定的方法进行整理、加工,并使之系统化的各种目录、索引和文摘。因此,二次信息资源仅是对一次信息资源进行系统化的压缩,无新的信息产生,具有汇集性、检索性的特点。它的重要性在于为查找一次信息资源提供了线索,是打开一次信息资源宝库的钥匙,可节省人们查找信息的时间。

3. 三次信息资源

三次信息资源是根据一定的目的和需求,在大量利用一、二次信息资源的基础上,对有关信息进行综合、分析、提炼、重组而生成的再生信息资源。如各种教科书、技术书、参考工具书、综述、年鉴、手册、指南等都属三次信息资源的范畴。三次信息资源具有综合性高、针对性强、系统性好、知识面广的特点,有较高的使用价值,能直接提供参考、借鉴和利用。

(二)按信息的载体形态分

1. 文献信息资源

文献信息资源是指以文献为载体的信息资源。文献信息资源依据其记录方式和载体材料又分为刻写型、印刷型、缩微型、视听型、机读型五大类。这五大类还可以进一步细分,如刻写型文献信息资源可分为手稿、日记、信件、原始档案、碑刻等;印刷型文献信息资源可分为图书、报刊、特种文献资料、档案、图片、乐谱等;缩微型文献信息资源可分为缩微胶片、缩微胶卷、缩微卡片等;视听型文献信息资源可分为唱片、录音带、录像带、电影胶卷、胶片、幻灯片等;机读型文献信息资源可分为磁盘、光盘等。

2. 实物信息资源

实物信息资源是指以实物为载体的信息资源。依据实物的人工与天然特性又可将实物信息资源分为:以自然实物为载体的天然实物信息资源(如土地、河流、植物、动物等)和以人工实物为载体的人工实物信息资源(如产品、样品、样机、模型、雕塑等)。

3. 网络信息资源

网络信息资源是指从计算机技术、通信技术、多媒体技术相互融合而形成的网络上可查找到的信息资源。按人类信息交流的方式,网络信息资源可分为以下三种:

(1)非正式出版信息。非正式出版信息是指流动性、随意性较强的,信息量大、信息质量难以保证和控制的动态性信息。如电子邮件、专题讨论小组和论坛、电子会议、电子布告板等工具上的信息。

(2)半正式出版信息。半正式出版信息又称“灰色”信息。指受到一定的产权保护但没有纳入正式出版信息系统中的信息。如各种学术团体、教育机构、企业、商业部门、国际组织、政府机构、行业协会等单位介绍、宣传自己或其产品的描述性信息。

(3)正式出版信息。正式出版信息是指受到一定的产权保护,信息质量可靠,利用率较高的知识性、分析性信息,用户一般可通过万维网查询到。如各种网络数据库、电子期刊、电子图书、电子报纸等。

信息资源是一个发展着的有机体,信息资源的类型也不是一成不变的,而是动态发展的。随着科学技术的发展,新的信息资源类型将不断涌现。信息资源类型体系应及时吸纳、涵

盖这些新兴类型。另外,随着信息资源内涵与外延的深化、拓展,信息资源的分类标准与分类方法也可能发生变化,信息资源类型体系亦应及时地予以调整,以保持信息资源类型与其定义的一致性。

第三节 信息检索

一、信息检索

信息检索是收集、组织、存储一定范畴的信息集合,并供用户按需要查询信息的方法和过程。广义的信息检索概念包括信息存储和信息检索两个过程;狭义的信息检索是从用户的角度来理解,仅指从已经存储的具有检索功能的信息集合中查询所需信息的过程。

二、信息检索系统

信息检索系统是指拥有一定的存储、检索技术装备,存储经过加工的各类信息,并能为用户提供检索服务的工作系统。检索系统由下列要素构成:①信息数据库;②存储、检索信息的装备;③存储、检索信息的方法;④系统工作人员;⑤信息用户。因而,信息检索系统具有吸收信息、加工信息、存储信息和检索信息等功能。信息检索系统按使用的技术手段可分为手工检索系统、机械检索系统和计算机检索系统。目前,常用的是手工检索系统和计算机检索系统。

(一)手工检索系统

又称传统检索系统,是用人工查找信息的检索系统。其主要类型有各种书本式的目录、题录、文摘和各种参考工具书等。检索人员可与之直接“对话”,具有方便、灵活、判断准确、可随时根据需求修改检索策略、查准率高等特点。但由于全凭人的手工操作,检索速度受到限制,也不便于实现多元概念的检索。

(二)计算机检索系统

又称现代化检索系统,是用计算机技术、电子技术、远程通讯技术、光盘技术、网络技术 etc 构成的存储和检索信息的检索系统。存储时,将大量的各种信息以一定的格式输入到系统中,加工处理成可供检索的数据库。检索时,将符合检索需求的提问式输入计算机,在选定的数据库中进行匹配运算,然后将符合提问式的检索结果按要求的格式输出。主要特点是:①检索速度快,能大大提高检索效率,节省人力和时间;②采用灵活的逻辑运算和后组式组配方式,便于进行多元概念检索;③能提供远程检索。

计算机检索系统,按使用的设备和采用的通讯手段,可分为联机检索系统、光盘检索系

统和网络检索系统。联机检索系统主要由系统中心计算机和数据库、通讯设备、检索终端等组成,能进行实时检索,具有灵活、不受地域限制等优点,但检索费用较高。光盘检索系统主要由光盘数据库、光盘驱动器、计算机等组成,具有易学易用、检索费用低的优点。根据使用的通讯设备,又可分为单机光盘检索系统和网络光盘检索系统。网络检索系统是将若干计算机检索系统用通讯线路联结以实现资源共享的有机体,是现代通讯技术、网络技术和计算机技术结合并高度发展的产物,它使各大型计算机信息系统变成网络中的一个节点,每个节点又可联结很多终端设备,依靠通讯线路把每个节点联结起来,形成纵横交错、相互利用的信息检索网络。

三、信息检索语言

(一)检索语言

检索语言又称标引语言、索引语言、概念标识系统等,是信息检索系统存储和检索信息时共同使用的一种约定性语言,以达到信息存储和检索的一致性,是标引人员和检索人员之间的桥梁。检索语言是人与检索系统对话的基础,用途是描述文献特征和表达检索提问,目的是为了提提高检索效率。

当存储信息时,标引人员对文献内容进行分析,概括分析出若干能代表文献外表特征和内容特征的词语,并赋予一定的标识,如题名、著者、主题词等,作为存储与检索的依据,然后纳入到检索系统中。

当检索信息时,检索人员首先要对检索课题进行分析,同样形成若干能代表信息需求的词语。然后通过检索系统在数据库中匹配具有同样词语和标识的文献,找到自己所需的信息。

上述即为检索语言工作的过程。不同的检索语言构成不同的标识和索引系统,给用户提供不同的检索点和检索途径。

(二)检索语言的类型

检索语言可分为描述文献外表特征的语言和内容特征的语言。

1. 外表特征检索语言

外表特征检索语言是以文献外表特征作为文献存储的标识和文献检索提问的出发点而设计的索引语言。主要检索标识有:

(1)题名。以文献的名称(论文题名、图书书名、期刊刊名、网站名称等)作为检索标识,如题名索引、引用期刊一览表等。

(2)著者。以文献署名的著者、译者、编者的姓名,以及机关团体名称等作为检索标识,如著者索引。

(3)文献序号。以文献特有的序号作为检索标识,如专利号索引。

(4)引文。将文献所引用的参考文献的著者、题名、来源出版物抽取出来进行标引,如美国《科学引文索引》的引文索引。

(5)全文。从文献的全部内容中自动抽取、查找,是目前网上各类搜索引擎使用最多的方

法。

2. 内容特征检索语言

描述文献内容特征的语言是检索语言的核心,按其构成原理可分为三类,即分类检索语言、主题检索语言和代码检索语言。

(1)分类检索语言。按照学科范畴及知识之间的关系列出类目,并用数字、字母、符号对类目进行标识的一种语言体系,也称分类法。使用这种检索语言建立的信息检索系统可以反映知识的从属、派生、重合、交叉、并列等关系,较好地体现学科的系统性,便于检索人员从学科或专业途径查找文献,并根据自己的需要扩大或缩小检索范围。目前常用的分类法有《中国图书资料分类法》、《美国国会图书馆分类法》(Library of Congress Classification,简称LCC)、《杜威十进图书分类法》(Dewey Decimal Classification,简称DDC)、《国际专利分类表》(International Patent Classification,简称IPC)等。

《中国图书资料分类法》将全部知识分为五大部分,并在此基础上分为22个基本大类,每个大类再逐级细分,层层展开,形成严格有序的体系分类表。

(2)主题检索语言。主题检索语言由主题词构成,即将自然语言中的名词、术语经过规范化后直接作为检索标识,按字母顺序排列标识,通过参照系统揭示主题概念之间的关系,也称主题法。包括先组式的标题词语言、后组式叙词语言及关键词语言。主题检索语言表达的概念比较准确,具有较好的灵活性和专指性,不同的检索系统、不同的专业领域可以有各自的主题词表。

1)标题词(Title Word)语言:它是用规范化的科技名词标引文献的主题概念,是受控式、先组式语言(即在检索前已经将概念之间的关系组配好)。具有较好的通用性、直接性和专指性,但灵活性较差。常用的标题词表有《美国国会图书馆的标题词表》(Library of Congress Subject)、美国《化学文摘》的普通主题索引和化学物质索引,它们是国外检索工具中使用标题词语言的典型。

2)叙词(Descriptor)语言:是从文献中优选出来经严格规范化的词,既表达文献内容特征,词间又有严密的语义关系。叙词语言最主要的特性是组配性,检索人员在检索时,只要根据检索的需要,临时从词表中选出相应的叙词,按照组配规则,任意扩大或缩小检索范围,极大地体现了叙词语言检索的灵活性。叙词的组配属概念组配,不是单纯的字面组配,概念组配是合乎逻辑关系的组配,美国《医学索引》的主题部分就是采用叙词语言的典型。叙词语言是多种检索语言的原理和方法的综合。它具有较优越的检索功能,是发展最快,应用最广的检索语言。

3)关键词(Keyword)语言:是直接选用文献中具有实质性意义的词,属非规范化词,关键词之间没有语法关系,易因概念混淆造成漏检或误检,适用于计算机检索。如美国《化学文摘》每期的关键词索引、美国《生物学文摘》的主题索引都属于关键词语言。

(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。就事物的某一方面特征,用某种代码系统来加以标引和排列,目前主要应用于化学领域。例如,化合物的分子式索引系统、环状化合物的环系索引等。

四、检索效率

检索效率是反映检索系统实施检索的有效程度和检索系统功能的指标。克兰弗登提出了6项评价检索系统性能的指标,即收录范围、查全率、查准率、系统响应时间、用户费用和输出形式。其中,查全率和查准率是两个常用的指标。

(一)查全率

查全率是评价系统检出相关文献的能力,为检出相关文献数与检索系统中相关文献总数的比值。

即:查全率 = 检出的相关文献数 / 检索系统中相关文献总数 $\times 100\%$ 。

(二)查准率

查准率是评价系统拒绝不相关文献的能力,为检出的相关文献数与检出文献总数的比值。

即:查准率 = 检出的相关文献数 / 检出的文献总数 $\times 100\%$ 。

由此可见,查全率和查准率之间表现为逆相关关系。如在检索时使用的检索词泛指性强,检出的文献多,查全率就会提高,但误检率也同时增大,因而查准率降低。如检索词专指性强,查准的文献多,查准率就会提高,但漏检率也同时增大,因而查全率降低。所以,要达到较好的检索效果必须两者兼顾,不能单纯追求某一指标,同时要根据检索的目的和要求,选择恰当的检索词。据有关调查显示,课题的查全率达到60%~70%,查准率达到40%~50%时,检索效果较好。

(三)如何提高查全率和查准率

1. 如何提高查全率

- (1)主题词和关键词检索相结合;
- (2)多用同义词进行 or 逻辑运算;
- (3)主题词检索时要进行扩展检索,并使用全部副主题词组配;
- (4)分类检索时要进行扩展和全部复分检索。

2. 如何提高查准率

- (1)用主题词和副主题词进行组配检索;
- (2)采用字段限定检索;
- (3)用 and 进行逻辑运算检索;
- (4)用出版年、期刊名称、文献类型等限定检索;
- (5)加权检索(限定在主要概念主题词字段检索);
- (6)强制检索(给检索词加引号)。