

Z H O N G Y I Y A O X I N X I J I A N S U O Z H I N A N

中医药信息 检索指南

主编 赖谦凯

文献检索的产生和发展文献检索的原理和基本知识文献检索的方法途径与步骤
文献检索的意义和作用医学文献检索工具信息意识与信息利用
中文数据库外文检索工具外文全文数据库
国内外开放存取资源及服务平台外文文献服务平台中国科技论文在线
古代中医药文献检索现代中医药文献检索

河南人民出版社

ZHONGYIYAOXINXIJIANSUOZHINAN

中医药信息 检索指南

主 编 赖谦凯
副主编 李 丹
张晓利

河南人民出版社

图书在版编目 (CIP) 数据

中医药信息检索指南/赖谦凯主编. - 郑州:河南
人民出版社,2010.5
ISBN 978-7-215-07187-2

I. ①中… II. ①赖… III. ①中国医药学-情报检索
IV. ①G252.7

中国版本图书馆 CIP 数据核字(2010)第 060041 号

河南人民出版社出版发行

(地址:郑州市经五路 66 号 邮政编码:450002 电话:65788050)

新华书店经销 河南省瑞光印务股份有限公司印刷

开本 787 毫米×1092 毫米 1/16 印张 20.75

字数 300 千字

2010 年 5 月第 1 版 2010 年 5 月第 1 次印刷

定价:50.00 元

目 录

第一章 总论	1
第一节 信息、知识、情报的概念	1
第二节 文献检索的产生和发展	5
第三节 文献检索的原理和基本知识	9
第四节 文献检索的方法、途径与步骤	13
第五节 文献检索的意义和作用	17
第六节 医学文献	20
第七节 医学文献检索工具	27
第八节 信息意识与信息利用	30
第二章 中文数据库	32
第一节 中国生物医学文献数据库(CBM)	32
第二节 维普期刊全文数据库	54
第三节 万方数据	61
第四节 中国知网(CNKI)数据库	69
第三章 外文检索工具	87
第一节 美国《工程索引》	87
第二节 美国《科技会议录索引》	97
第三节 《科学引文索引》	99

第四章 外文全文数据库	115
第一节 Elsevier 数据库	116
第二节 EBSCOhost 数据库	120
第三节 SpringerLink 数据库	131
第四节 OVID 系统数据库	135
第五节 PQDD 博硕士论文数据库	141
第五章 国内外开放存取资源及服务平台	147
第一节 PubMed	148
第二节 GoPubMed	158
第三节 外文文献服务平台——FMJS	164
第四节 中国科技论文在线	183
第六章 古代中医药文献检索	199
第一节 古代中医药文献概述	199
第二节 中医药古代文献分类检索	201
第三节 中医药文献专题检索	243
第七章 现代中医药文献检索	267
第一节 现代中医药文献概述	267
第二节 中医药书刊、期刊检索	270
第三节 中医药论文检索	278
第八章 中医药情报信息研究	294
第一节 科技查新	294
第二节 情报调研	303
第三节 情报分析方法	316
第四节 情报性文献写作	321
后 记	330

第一章 总 论

人类正以前所未有的速度迈入信息社会,信息作为继材料、能源之后的第三种重要资源,已成为支撑当今社会经济发展的三大支柱之一。高速度地传播信息,及时地占有信息,准确地加工、提炼信息,有效地利用信息,在改造自然、推动社会进步的进程中实现信息的价值,是信息社会的本质特征。不断地更新知识,快捷、有效地掌握和利用信息,是对 21 世纪人才的基本要求。

随着医学科技的发展,医疗信息在人类社会发展中有着重要的作用,专业知识信息的获取、利用,对于每个从事科研、教学和临床工作的医学工作者都是十分重要的,而充分利用文献检索工具,能够全面、及时、准确地得到所需的知识信息,在工作、学习中达到事半功倍的效果。

第一节 信息、知识、情报的概念

一、信息 (information)

信息广泛存在于自然界、生物界和人类社会。物质、能量、信息是构成客观世界的三个基本要素。信息是一种普遍存在的现象,是信号、消息,是事物存在和运动状态的反映,是认识的基础。信息是一个十分普通而复杂的概念,人们对信息有着不同的认识和理解。

N. 维纳在 1948 年发表的《控制论——动物与机器中的通讯与控制问题》中指出:信息既不是物质,又不是能量,信息就是信息。信息就是我们适应外部世界并且使这种适应为外部世界所感知的过程中,同外部世界进行交换的内容的名称。信息是“生物以及具有自动控制系统的机械系统,

通过感觉器官和外界交换的一切内容”。哲学上认为,信息是事物本质、特征和运动规律的反映。不同的事物有不同的本质、特征和运动规律,人们就是通过事物发出的信息来认识该事物,并以此区别于其他事物的。

信息是指一切音信,是客观世界一切事物的存在和变化状况与人的认识能力的交换。信息的基本特征是具有广泛性、可传递性、可分享性、可选择性、知识性、效用性、可存储积累性、衰竭性等。信息和每个人息息相关,人类的任何社会实践活动都建立在获取、存储、加工、利用、传播信息之上,并自觉地对获取、存储、加工、传播的信息做出有选择的反映。当今社会人类及个人事业的成败原因,均可以归纳为掌握和应用信息的差异。

信息这个词在我国唐朝时期就出现了,唐朝李中在《春款怀故人》中有“梦际美人沉信息,目穿长路倚楼台”的诗句。宋朝陈亮在《梅花》中有“欲传春信息,不怕雪埋藏”的诗句。两诗都有信息一词,其含义大致是迹象、征兆与消息,运用时都带有自然语言固有的随意与模糊色彩,不妨碍沟通与理解。书目文献出版社出版的《图书馆学情报学辞典》一书中解释说:“从哲学的观点看,信息是物质的一种带有普遍性的关系属性,是物质存在方式及运动规律、特点的外在表现;从通讯的角度考虑,可认为是生物体或具有一定功能的机器通过感觉器官或相应的设备同外界交换的内容总称。”因此,信息这一术语适用范围很广,如树的年轮表现了其生长的年龄与发展的信息;海水的涨落表现了月亮围绕地球运行的信息;人的体温的升高表现了人体患病的信息。所以信息已成为自然、社会、生命、思维等各个领域内普遍化的术语。由于事物在不断地发展变化,它的运动状态及其表现出的特征亦在不断地变化着,而同属于这种运动状态和变化的规律的信息,也同时在不断产生。不同的事物和其不同的矛盾运动状态就会产生不同的信息。自然界本身的运动变化无时不在产生大量信息;人对自然界的作用与反作用也无时不在产生大量的信息;人类本身及其在医药领域的活动及其变化,人类与疾病作斗争的各种运动过程,也就不断地产生各种医药信息。因此,信息被认为是无所不在、无所不有的在整个宇宙中普遍存在的东西。然而人们对信息的发现和认识受各个时代生产力和科学技术发展的水平及认识能力的影响和制约。因此,可以说,人类社会发展的历史就是人类不断认识信息、获取信息、掌握信息、传递信息、生产信息,并用这些信息为人类服务、改造客观世界、也改造主观世界的历史。但是随着社会的不断发展,信息的产生和积累愈来愈多、愈来愈复杂,人们需要获得、传递、掌握的信息也越来

越多,用来解决问题的范围也日益广泛,于是,一个信息时代、信息社会就出现在人们的面前,使人们感到,自己处在信息的包围之中,如不及时处理、挑选并采用自己所需要的信息,就有可能产生被社会淘汰的危险。上世纪 80 年代以来,我国出现的“四热”中就有“信息热”,传播媒体中有商品信息、经济信息、购物信息、人才信息、信息世界、信息导报、医学信息报,等等。信息这个词已成了高频而广泛流行和使用的词汇了。

二、知识(knowledge)

知识在《现代汉语词典》中的解释为:“人们在改造世界的实践中所获得的认识和经验的总和。”在《图书馆情报学词典》中的解释为:“人类对客观事物的认识、实践经验的总结。解决问题的方法属于认识的范畴。”根据毛泽东同志《实践论》的观点,人们在改造大自然和改造社会的实践过程中,通过自己的感觉器官,不断地感触世界,了解世界,并不断地对感觉到的事物进行总结、加工、提炼、系统化,上升为理论,再反过来将这些知识用来指导实践,又获得新的知识。依次下去,反复循环,使人们的认识不断提高、不断深化、不断达到新的高度。这种认识,可算是人类的第二资源,它可以指导和推动人们开发第一资源。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。所谓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起来的经验的结晶,它是人们与生老病死作斗争的有力武器。

三、情报(information)

情报是关于某种情况的消息和报告,习惯上指的是“战时关于敌情之报告”,由此发展为“推动用户当前事业的信息选择、传递”。笔者对此则解释为运用一定的媒体(载体)越过空间和时间传递给特定用户,解决其在科研、生产中所需要的特定的知识或信息就叫做情报。一般而言,情报应包括三个要素:一是知识或信息;二是要经过传递;三是要经过用户实用产生效益。例如一个科研机构研制成功一种效果很好的癌症晚期止痛药物,那么有关这个药物的知识和信息通过某种载体从这个科研机构传给某个医院需要这种药物来为癌症病人止痛的病人和医生,应用后并产生了效果,那么这个药物的知识或信息就叫做情报。而对于这个医院里不需要这个药物的医

生和病人就不是情报,只能说,他们了解后,获得了这种药物的知识或信息。因为这些药物他们既不需要,又不能在他们这里产生效益。正如钱学森同志所说,“情报是知识的激活”,是一种普遍存在的社会现象。自从有了人类社会,就有了情报。随着社会的进步,阶级、社会、贸易竞争的出现,人们对于情报的需要愈来愈多,情报的内容也愈来愈广泛,交流的手段和方式也日益多样化,特别是现代科学技术日新月异地发展,市场竞争日趋加剧,除了军事情报、经济情报、政治情报和科技情报外,还有了商品情报和市场情报等,情报已渗透到各个行业各个领域,成为经济、政治、军事、文化、科学技术向前发展的支撑条件。一些发达国家,比如日本把情报作为富民强国的向导,使其经济得到了高速发展,成为世界经济大国。一些发展中国家把情报作为发展的催化剂;各个部门的领导依靠情报来作出正确的决策,广大科技工作者不断猎取情报,大大加快了科学技术的发展。今天情报在全世界的作用显得越来越突出了。

四、文献(literature, document)

文献是人类长期从事生产活动和科学技术活动以及社会交往的真实记录,是具有一定历史文物价值的珍贵资料,是人类物质文明和精神文明不断发展的产物,是精神财富的重要组成部分。它记录了人类历史长河中科学技术发展和人类社会活动所达到的成就和水平,凝结着数千年来人们的辛勤劳动和智慧,积累着各种对后人有用的事实、数据、理论、方法,记载着前人成功的经验和失败的教训,反映各个时代各种社会环境下科学技术和人类社会进步所达到的水平,预示着将来发展的趋势和方向。人类在漫长的生产实践、科学实践、社会实践中逐步认识客观世界,就产生了大量有用的知识,这些知识或信息则通过一定的形式体现出来,如用文字、图像、声频、视频等反映出来。为了把这些知识积累起来,便于传播,被后人借鉴,发挥出巨大的作用,人们就设法将这些知识或信息用一定的符号、文字、图像等记录在一定的物质形态上,这就形成了文献。那么,在人类出现以后,就面临着生老病死各种问题,就有了与生老病死作斗争的经验,为把这些经验保存下来,为后人所借鉴,就将其记录在一定的载体上,这就是医学文献。换言之,凡是人类所获得的医学知识,以一定的方式,记载在载体(骨、泥板、石、金、皮、竹木片、绵帛、纸张,发展到现在,则有了胶片、胶卷、塑料、磁带、磁盘、磁鼓、光盘等)上的每一件记录,统称为医学文献。简言之,将医

学知识记录在特定的载体上的结合体称为医学文献。载体(媒介)、知识(信息)是文献的两个基本要素,记录是文献的基本特点,三者相辅相成,组成文献这个整体,缺其中之一都不能成为文献。医学知识或信息是医学文献的实质内容,载体是医学文献的外部形态,记录是两者之间的联系物。不通过记录,知识就不可能记载在物质载体上,就不可能把知识和载体结合起来,就不可能成为文献,就无法保存下来,人类所获得的知识就会失传,后人在生产和科研上就可能走前人的老路,人类社会的发展就可能要大大延缓。过去不少人把知识或信息与文献的概念搞混了,甚至在一些书里出现了“把文献记录在载体上”的话,这是不妥的,因为文献本身已是将知识记录在物质载体上的集合体了,它不应当也不可能再将自己记录到载体上了。

五、文献与信息、知识、情报的关系

从前述我们可以清楚地看出,文献与信息、知识、情报之间有着极为密切的关系。信息知识、情报是抽象而又十分复杂的概念,它必须通过一定的文字、符号、图像、音像等加以物化,形成一定的物质形态,才能表现出来,才能看得见、摸得着。文献既不同于纯粹的物质载体,也不同于纯粹的知识、情报、信息,然而文献不能脱离物质载体或知识、情报、信息。文献是精神信息与一定的物体的结合,是重要的知识源、情报信息源,它是知识、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;信息是知识的重要组成部分,但不是全部,只有提高了的、深化了的系统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,被激活了的那一部分就是情报。知识、情报、信息的主要部分包含在文献之中,成为构成文献的三要素。医学情报是医学信息通过思维、加工、提炼、重新组合并系统化了的的知识。这些都必须固定在一定的物质载体上、形成文献后才能进行传递,才能被人们利用。

第二节 文献检索的产生和发展

一、文献检索(document retrieval)

检索就是查找的意思,所谓文献检索就是查找文献,就是将文献按其外

表特征或内容特征,按照一定的方式编排并存储在一定的物质载体上,并利用相应的方法、途径或手段从中找出符合读者特定需要的文献的全过程,属于情报信息检索的范围。它有两个方面的内容:一是文献的存储,就是将大量分散的文献收集起来,经过加工、整理,使之有序化和系统化,成为有查寻功能的检索工具,这项工作一般由图书情报工作人员去完成;二是文献检索,就是利用已编制好的检索工具,按读者的研究课题所需要的特定范围,将文献检索出来,这项工作多数是科技人员自己去做,也可以请图书情报工作人员去做。目前在国内的机检工作,一般由图书情报人员去完成,科技人员自己动手的则甚少。总之,文献检索一是放进去,二是取出来,放进去是手段,取出来才是目的;只有科学地放进去,才能顺利地取出来,放进去多少,就能取出来多少,放进去是基础,是取出来的前提,输入(放进去)一次,可以反复输出(取出来)利用多次,即平常所说的一次存储、多次检索。

我们平常所说的检索,不包括前面的存储部分,而限于后一部分文献的检索,它按不同的加工深度,分为题录检索、文摘检索、全文检索;按检索的设备分为手工检索和机械检索,机械检索又可分为脱机检索、联机检索、光盘检索等等。

有些人往往把文献检索与情报信息检索看做是一回事,其实这二者是有区别的。情报信息检索的对象包括了数据检索(data retrieval)、事实检索(fact retrieval)和文献检索(document retrieval)三种检索。而文献检索的对象只限于文献,所以情报信息检索的范围比文献检索要广,文献检索只是情报信息检索的一部分。

二、文献检索的产生与发展

文献检索的产生与文献的产生和发展息息相关,是在人类社会不断发展,保存积累了一定数量的文献以后。在公元前15~公元前14世纪我国的殷商奴隶社会时期,史官们把记录当时经济、政治和文化的文献集中保藏。为了查找和使用方便,按照一定次序加以排列,并编制出相对应的数码,在这个过程中不断地总结,逐渐得出了固定的方式和规律,另外编成单据,这就产生了简单的著录文献的目录。这是我国古代文献检索的萌芽时期。在一些甲骨上刻有“入”、“示”和一些数码,这就是主管保存文献的人所做的记号,而这些记号和数码与另外的简单单据或目录是相对应的,这都表示了检索工作的实际意义,包含着检索工作的雏形,代表着我国古代文献

检索工作的起源。

春秋时期(公元前 770 ~ 前 476 年)奴隶社会已到末期,文献的积累和保存更丰富,检索工具的编制和利用也有了一定的发展。从孔子学派校书的大序小序,到战国秦汉诸子百家著书的自序和校书的叙录,是我国古代目录中提要或解题从发生到发展的形成过程,那些大序、小序在当时实际上起着提要(今天的文摘)目录的作用。公元前 1 世纪末,由于经济、政治、军事、文化发展的需要,我国第一部系统目录——刘向的《别录》产生,这是综合性的图书分类目录,《七略》是刘向之子刘歆编成的简略系统目录,提供了较丰富的图书内容的说明,便于参考使用。西晋时期编成的《晋中经簿》是我国第一部官修目录,在分类和著录上都做了适当的改革。东汉初年班固编成的《汉书·艺文志》,第一次开辟了纪传体史书内“艺文志”的做法,成了官修目录以外的一个新品种。公元 5 ~ 6 世纪,王俭编的《七志》、阮孝绪的《七录》是这一时期有影响的检索工具。公元 4 ~ 5 世纪,陆澄的《法论目录》和曾佑编的《弘明集目录》等,在佛经目录参考之外又包括了论文目录,使目录兼有了索引的作用,这是我国古代索引的胚胎。公元 618 ~ 1380 年,是我国封建社会的发展高峰时期,文献检索有了较大的发展,以政府的官修目录为基础,编成了《经籍志》和《艺文志》,这种目录的分类和编成提要等方法的发展变化又对私人藏书目录起了示范和促进作用。专科目录又在上述目录的影响和制约下综合利用索引与提要向更专、更深的方向发展。唐代有《群书四部录》和《古今书录》,宋代有《崇文总目》和《中兴阁书目》,并且还出现了指导阅读的书目,《杂钞》是敦煌出土最早的推荐性书目。南宋和元代,是我国文献检索达到兴盛繁荣并开始衰落的时期,出现了郑樵的《通志·艺文略》和马端临的《文献通考经籍考》目录巨著。清朝的《四库全书总目》是我国封建社会最大的一部官办的检索工具,共 200 卷,收录图书 10231 种、177003 卷,全部目录分经、史、子、集四部,44 个大类,还细分为 67 个子目类,对我国文献检索在理论上产生了较大的影响。在此期间,我国的目录学家章学诚的目录学思想批判地继承了目录学方法论,达到了一个较高水平。我国古代文献检索的发展与成就远远超过了同时代世界其他各文明国家,不仅有系统分类目录的优良传统,还有了互著别裁分类著录的辅助方法。章学诚在《校雠通义》内既阐述了系统目录中图书分类著录的重要意义,又精辟地发挥了互著别裁在分类著录中的辅助作用,这种互著别裁实际上就是我们今天检索工具中的参照系统。鸦片战争以后,值得提出的是

梁启超的《西学书目表》，在我国类分图书方面最先提出按自然科学、社会科学、综合性图书三大部类分类，不久，康有为编辑了《日本书目志》介绍日本明治维新以来的书刊，还有《古越藏书楼书目》开创了我国近代图书馆目录之先河。五四运动以后，新文化运动发展，不仅书目有了很大的发展，而且索引也有了空前的发展，出现“引得”（索引）数十种之多，写下了我国文献检索的新篇章。在此期间出现了文献检索是揭示报章文献与人们对文献特定需要之间的矛盾这一概念，这已达到了对文献检索本质的认识。

20 世纪 50 年代，我国的科学事业获得了迅猛的发展，反映这种发展的文献随之成倍增长；另一方面，由于科学技术发展的需要，每年引进了大批国外的书刊。科学技术发展使得学科越分越细、分支学科越来越多，人们需要查阅文献的范围变得越来越狭窄，于是，巨大的文献资料海洋与人们需要的特定而狭窄的学科范围的文献之间的矛盾也愈来愈尖锐化。为了很好地解决这一矛盾，就需要先进的文献检索方法和理论，以及先进的检索工具及现代化的技术设备，这样就促进了文献检索以前所未有的速度向前发展。50 年代后期，各种书目、文摘、索引不断涌现，出版了相当数量的检索刊物。与此同时还引进了国外一些主要的检索刊物，如美国的《医学索引》（*Index Medicus*）、《生物学文摘》（*Biological Abstracts*）、《化学文摘》（*Chemical Abstracts*）和荷兰的《医学文摘》（*Excerpta Medica*）等等。但是，有关资料表明，我国会文献检索的科技人员极少，95% 以上的都不会查文献，以致造成了大量的重复劳动，浪费了大量的人力、物力、财力。为了充分地发挥这些检索工具的作用，就需要科技人员掌握它们的编排原理和查找方法，以便从文献的海洋中获得所需要的对口文献，从中猎取有关的情报信息，以推动科学技术的发展。因此，在这个时期有不少图书情报机构，采取了一些有效措施，不断地向科技人员宣传并讲授文献检索知识，在有的大学已开始给教师和高年级学生讲授文献检索知识，并已取得了一些经验。为巩固和扩大成果，1982 年由南京医学院图书馆馆长吴观国牵头，组织全国 10 所重点高等医药院校中具有多年文献检索工作实践经验又有较高理论水平的老同志，编写出了我国第一部医学文献检索教材《医学文献检索》，并在次年 4~5 月举办了第一期全国高等医药院校文献检索教学师资培训班，该教材在这个班试用后获得较好的效果。1984 年暑假又在中国医科大学举办了第二期师资培训班，总共培养了 150 余名具有系统医学文献检索理论知识和方法的教师，他们在以后的文检课教学中成了骨干。与此同时，各级领导和有关

人士也看到了对学生和科技人员进行文献检索知识教育的重要性,特别是国家教委,从1984年到1992年的8年间就全国高等学校开设文献检索课的问题,先后下发了三个文件。这样,文献检索作为一门课程不仅得到了国家的确认,而且正式确立了它在高等学校课程中的地位,学校文献检索教育得到了很大的发展,使高等学校的教师、学生文献检索的能力大大加强了,同时对社会各影响也较大。为了适应医学文献检索教学发展的需要,经过一年的筹备,于1990年10月在武汉召开了全国医学文献检索教学研究会成立大会暨第一届学术研讨会,这标志着我国医学文献检索教学进入了新阶段。从此,我国的医学文献检索及其教学获得了更加稳定的发展,产生了巨大的经济效益和社会效益。

20世纪60年代后期,美国将电子计算机应用于文献检索,揭开了文献的机械检索的新篇章。美国国家医学图书馆研制成了“医学文献分析与检索系统”(MEDLARS),是将计算机应用于文献检索最早最好的典范,后来发展成为全美国直到国际上的联机检索系统,在医学文献检索中起到了难以估量的作用,我国80年代初开始将计算机应用于医学文献检索试验研究,80年代后期才进入实用阶段,90年代以来有了很大的发展。今天我们不仅在手工检索系统方面有了空前的发展,而且在机械检索方面也有了相当的规模,形成了一个手工检索和机械检索并行发展、互相支持、互相补充、互相促进的局面。不仅有了联机检索和光盘检索,而且已经有某些地区或某些系统的联网检索,也有了国际联机检索。不仅有了大批训练有素的医学文献检索的专业人员,而且还有了能从事各个层次的文献检索教学的师资队伍。医学科技人员的文献检索能力大大提高了,盲目地搞科研、重复别人的科研课题的现象已有所减少,科研速度加快了,周期缩短了,文献检索正在沿着正确的轨道不断地前进!

第三节 文献检索的原理和基本知识

一、文献检索的原理

文献检索的基本原理就是检索者将检索提问的标识符与存储在检索工具中的文献特征标识进行比较,二者若能取得一致,则可以将检索工具中符合提问要求的文献提取出来。文献存储和检索是不可分割的两个过程。存

储过程主要是对文献进行标引,形成文献的外表和内容特征的标识,为文献信息检索提供有章可循的途径的过程。文献的外表特征是指标题、著者、来源、卷期、页码、年月、文种等项目;内容特征指文献论述的主题。文献存储前,首先要对文献内容进行主题分析,把握所论述的中心内容,形成主题概念,然后选用特定的文献检索语言(如主题词、分类语言等)表达主题概念,也就是将主题概念转换成文献内容特征的标识(主题词、分类号和类目名称),最后将这些标识进行编排,输入文献检索系统。

文献检索过程是根据读者需求,先确定检索概念及其范围,然后选择一定的检索语言,并将检索概念转换成检索特征标识,按此到检索系统中查找文献线索,最后对其进行逐篇筛选,确定需要阅读的文献。

二、检索语言

检索语言是人们在信息检索领域中用来描述文献外表特征和内容特征及检索提问的一种专用语言,是检索中使用的标识系统,是由给定领域中一切可用来描述信息内容和信息需求的词汇、符号及其使用规则构成的供标引和检索使用的工具。在检索系统中起着语言保障的作用。它是标引人员和检索人员双方沟通的桥梁,保证标引用词和检索用词的一致。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。

(一) 外表标识检索语言

外表标识检索语言是描述文献外表特征的检索语言,是依据文献外表特征,以文献题名、著者、文献序号等作为文献存储的标识和文献检索提问的出发点设计的索引语言。主要有:

1. 文献题名索引系统

以文献上记载的书名、刊名、篇名等作为标识的字顺索引系统,如引用期刊一览表、篇名索引。

2. 著者索引系统

以文献上署名的著者、译者、编者等姓名或机关团体名称作为标识的字顺索引系统,如著者目录(索引)、专利权人索引。

3. 文献序号索引系统

以文献特有的序号作为标识的索引系统,如专利号索引、(科技)报告号索引、技术标准号索引。

4. 引文索引系统

是根据末尾所附参考文献引文索引系统及引用文献这一外表特征,揭示科学论文之间引证与被引证关系而编制的索引系统,如美国《科学引文索引》(*Science Citation Index*)。

(二) 内容标识检索语言

内容标识检索语言是描述文献内容特征的检索语言,按其构成原理可分为三类,即分类语言、主题语言和代码语言。

1. 分类语言

分类语言是将文献根据其所属的学科内容分门别类地系统化组织的一种方法。这种分类能够较好地体现学科的系统性,揭示知识的派生、隶属与平行关系,便于检索者从学科或专业途径查找文献资料,并根据自己的需要扩大或缩小检索范围。

国内外的图书分类法,如《美国国会图书馆分类法》(*Library of Congress Classification*, LC)、《杜威十进分类法》(*Dewey Decimal Classification*, DC 或 DDC)、《国际十进分类法》(*Universal Decimal Classification*, UDC)等。《中国图书馆分类法》和《中国图书资料分类法》(简称《中图法》)是我国使用最普遍的分类法。《中国图书资料分类法》和《中国图书馆分类法》在类目结构、分类原则等方面基本相同,只是前者类目等级设置更细,最多达到八级,主要用于各类文献资料,如期刊论文资料的分类。

《中图法》属于等级体系分类语言。它以知识分类为基础,结合文献的其他显著特征(如地区、时代、形式、体裁等),依照从总到分、从一般到特殊、从低级到高级、从简单到复杂的逻辑次序,层层划分,层层展开,形成严格有序的体系分类表。

2. 主题语言

主题语言是用主题词来标引和检索文献内容的检索语言,包括标题词语言、关键词语言、叙词语言等。

(1)标题词语言。标题词是指来自普通语言中那些比较定型的事物名称,而且经过规范化,用以表达文献主题内容的词、词组或短语。简而言之,标题词就是比较定型的事物名称。由标题词组成的索引标识系统,称为标题词语言。许多检索工具的主题索引都是用标题词语言编制的。如美国《化学文摘》普通主题索引及化学物质索引。

(2)关键词语言。关键词(key words)是直接从文献的篇名、摘要和正

文选出的具有实质意义并能代表文献主题内容的词、术语,由关键词组成的索引标识系统,称为关键词语言。它是一种未经优选和规范化的自然语言。如美国《化学文摘》(CA)每期的关键词索引、美国《生物学文摘》(BA)的主题索引等。

(3)叙词语言。叙词(descriptor)是指来自文献、用户及医学专家,并经过严格的多方面控制,用以表达文献主题或信息需求的单义词或代码。叙词语言是在吸取了多种检索语言优点的基础上形成的一种检索语言。叙词语言适用于计算机化的文献检索。

(4)代码语言。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,如美国《化学文摘》的分子式索引、环系索引等。

三、检索效率

评价检索效率的指标有多种,但较常用的有查全率和查准率。查全率是评价系统检出相关文献的能力,为检出的相关文献量与系统中相关文献总量的比值。查准率是评价系统拒绝不相关文献的能力,为检出的相关文献量与检出文献总量的比值。课题的查全率一般为60%~70%,查准率为40%~50%。不少文献信息专家还特地用数学方法和坐标图来证明文献查全率与查准率之间存在的互逆关系,如要使查全率超过70%,就只能靠降低查准率来实现;反之,要提高查准率就要降低查全率。任何一种检索系统要求其查全率和查准率都达到100%是不可能的,只能在允许的条件下按照读者的要求,在一定范围内选择最佳方案来提高查全率和查准率。

专家们用相关度来解释这种互逆关系规律。检索开始时,检索者总是找相关度大的文献,因为它的特征明显,查找方便、快速,但往下检索,相关度越低,判断相关性越加困难,容易误检,速度慢。可见初检时查全率低而查准率高,但随着检索文献数量的增加,查全率愈高,误检率愈大,查准率降低。

文献检索出来的结果,一般是文献的线索,而读者往往需要原始文献,这些文献有的本馆入藏了,有的没有入藏,要靠兄弟馆的帮助来满足需要。因此,从图书情报机构和读者的角度来考虑,一方面要设法提高文献检索的效率;另一方面要提高文献的利用率,充分发挥文献的作用。这就需要从以下几方面入手:

一是加强文献的宣传报道和对读者文献检索知识的教育。很多单位入藏了丰富的文献,可是利用率很低,有的长期无人问津。不是这些文献没有