



情报检索教程

贺志刚 李修波 / 主 编

QINGBAO
JIAN SUO
JIAO CHENG

地震出版社



128

G74-2-03
A201

情报检索教程

贺志刚 李修波 主编

地震出版社

2000

前 言

(代 序)

我们最初开设情报检索课是在 80 年代初,是响应前国家教委“在有条件的高校开设公共情报检索课”的指示,对我校研究生和高年级本科、专科生进行“情报检索”基本知识的教育。课程开设以增强在校大学生的情报意识,提高他们的情报检索技能,从而最终达到有效提高科研工作效率目的。

科学技术的迅猛发展,信息量的激增,使人们获取情报越来越困难;同时随着计算机技术和网络技术的快速发展,情报检索的新方式、新方法也逐渐“浮出了水面”。新的历史条件对情报检索课提出了新的要求。一方面是广大情报用户,特别是作为潜在情报用户的在校大学生,他们迫切需要学会最新情报检索工具的应用方法;学会去利用最先进的情报检索手段来获取情报的知识。与此相对应的另一面则是,情报检索教学用的教材内容陈旧落后,已不能满足现代化情报检索教学工作的需要。编制一本适合现代化需求,并受大学生欢迎的教材成为一件非常紧迫的事情。

我们在繁忙的教学工作中发现了许多问题,并利用自己的经验对教学内容进行了调整。为编制一本水平和质量都较好的教材,我们对教材的正式出版采取了非常谨慎的态度。我们在两个学期分别印制了教材试用本试用,同时广泛地征求师生及任课教师的意见,汇集所有良好的建议,最后才编成了此教程。

在此要对我们的几位编委表示衷心的感谢,他们在本书的编撰工作中付出了辛勤的劳动,工作之余加班加点,精心修改,为教材的顺利出版做出了贡献。各位编委对本教程的章节具体责任如下:第一章、第二章、第四章第三节、第七章、第九章、第十一章等由贺志刚同志编写;第三章、第四章第四节、第五章第三节、第八章、附录等由李修波同志编写。宋文彬、左玉琴、王勋鸿、宋西贵、马建春、牟鑫等同志分别编写部分章节,程蓓、隋云昌同志参与修改了计算机情报检索部分。韩子军研究馆员作为一位资深的图书馆学者,在百忙之中对本书的内容进行了细致的审校工作,我们特别表示感谢。

我们编制的这本教程系统介绍了情报检索的基本知识,讲述了书本式中外文参考工具书、检索工具书及其利用方法;计算机情报检索方法与技巧及因特网情报检索等内容。另外扼要地介绍了情报收集、整理与积累的基本方法。在编写过程中我们力求做到:突出重点,照顾一般,以求学生学习时能达到举一反三的目的;阐述内容简明直观,检索工具基本都附有检索实例;为适应时代需求,加强了计算

机检索与光盘检索内容的论述。

我们希望这本教材带给您的不但有快捷查找情报资料的愉悦,还能有科研水平的长足进步。但由于我们水平所限,编辑时间较为仓促,疏漏、不足与错误之处在所难免,敬请批评指正。

最后,谨将衷心的感谢送给为本书的正式出版给予无私帮助的人们!

编 者

2000 年 6 月

第一章 科技文献概论

1.1 知识、情报与文献

人类世界正处在知识量激增的时代。随着科技人员数量的增加和日益庞大的信息系统的出现,特别是全球计算机网络的不断扩展和完善,使人类进入了信息化的时代,一方面,人类的智力在更广阔的时间和空间进行探索,使得科技成果以前所未有的速度增长。据估计,现在世界每年发表的科技论文大约 500 万篇,出版的图书约 70 万种,申请的专利约 100 万件,此外还有大量的特种文献出版,尤其是随着 INTERNET 的发展,网上传播的信息非常庞大。情报信息的载体形式也更加多样化,如有印刷型、缩微型、机读型、声像型等。另一方面,知识的更新速度在加快,情报的有效期在缩短,如 10 年前发展起来的新兴工业技术,今天已有 30% 过时,电子技术、通信技术、计算机技术更新速度更快。有人估计,图书平均利用期限约为 10~20 年,研究报告为 10 年,学术论文为 5~7 年,期刊为 3~5 年,国家标准 5 年,产品样本 3~5 年。因此在人类进入信息社会的今天,为了跟上时代的步伐,为了快速获取自己所需要的最新知识,掌握情报检索的基本知识对一个科研工作者来说很有必要。

知识是人类对客观世界的物质形态与运动规律的认识,这种认识来源于客观事物在人脑中的反映。随着反映的深入,从现象到本质,从感性到理性,人的认识不断深化,逐步掌握了事物的客观规律。知识是人类文明发展的动力,人类利用知识就会产生出改造客观世界的无穷无尽的物质力量。作为个体的人来讲,他所需要利用的知识仅是客观知识存在中的某一特定部分。

情报是知识的传递部分。它的基本属性就是知识性和传递性。因此,情报具有特定的知识内容,而且它是以运动形式存在的,是在特定条件下对特定目的有用的那一部分知识,是知识的一部分。

知识必须依附于一定的载体,如存在于人的大脑中,或固化在一定的物质载体上。人类在漫长的社会生活、生产劳动和科学实践中积累了大量的知识,要借助于一定的载体(如:泥块、甲骨、竹简、绢帛、纸张、胶片、磁带、磁盘、光盘……)进行记录与传播。这种以文字、图形、符号、声频、视频等手段记录和传播知识的载体就是文献。我国国家标准将文献定义为:“文献是记录有知识的一切载体”。因为情报的本质是知识,它又要借助于一定的载体加以传播,所以,科技文献就是科技情报的载体,而科技情报是科技文献的内容。

1.2 情报交流与科技文献

科学研究是探索客观事物的本质规律的活动,是人类认识世界和能动地改造世界的过程。当今的科学研究必须以占有丰富的情报资料为基础。科技文献是记录科学研究成果的信息载体,是人类智慧的结晶,它积累了大量的事实、数据、理论、方法和科学假设,记载了无数成功的

经验与失败的教训。科技人员渴望得知前人所做过的一切,也渴望得知他所从事的研究领域中别人正在做什么,以及研究的进展情况。这就需要掌握有关的情报,需要占有与自己的研究有关的资料。同时,科技人员进行的研究也需要以文献的形式公布,进入情报交流。

1.2.1 情报的交流与传递

情报交流指的是人与人之间借助于他们使用的共同符号(语言、文字、图形、图像、代码等)所进行的知识信息的交流,它是人类社会的一种普遍现象,是人类社会的基本要求。

情报交流可分为非正式交流和正式交流。

非正式交流(直接交流)是指科技工作者自己完成的情报传递方式,是在一定范围内由科学家之间或学术团体之间直接进行的。例如,科技工作者就他们所从事的研究进行面对面的直接交谈,参加学术会议、学术沙龙,参观同行的实验室,科技成果展览,私人通信,交换手稿等。

正式交流(间接交流)是借助于科技文献检索系统,在情报人员参与下进行的情报传递。这在当代社会已成为情报交流的主体。

通过非正式交流来获取情报在科研中起着重要的作用。据 M. H. 哈尔伯特和 R. L. 阿克夫的研究资料,全部科学情报大约有 1/3 是通过非正式渠道传递的。非正式交流所获得的情报较正式渠道获得的情报要迅速得多,而且均为最新情报,可直接、迅速、生动地反映出科研活动的最新动向。特别是 INTERNET 的建立,为科技人员提供了通过电子邮件(E-mail)进行非正式交流的便利条件。对于科技人员而言,为了能有效地利用非正式渠道获取情报,必须具有敏锐的情报意识,随时捕捉网罗对自己有用的情报,消化并“引种”到自己的头脑中,以触发新思想和新观念的产生。

非正式交流的方式出现在正式交流之前,是早期的情报传递方式。那时,发明创造一般都是通过个人研究完成的。由于科学技术还不发达,科学家人数不多,发表的文献数量有限,科技情报的交流主要是通过非正式渠道进行的。但非正式交流有其不可避免的弱点:第一,适用的范围有限,往往只有少数人能有机会参加直接交流,对于刚刚进入科研领域的新手来说,这种机会更少;第二,缺乏有效的社会监督,没有经过科技界评价的情报,缺乏一定程度的可靠性;第三,不便于今后的加工整理和存储积累。

随着科学技术的发展,科研工作的协作性越来越重要,科学家个人自由研究的方式已不能适应科学技术发展的要求。到了 19 世纪中叶以后,各国都陆续成立了不同的学术团体。例如:美国 1852 年成立了“美国土木工程师学会(American Society of Civil Engineers)”;1880 年成立了“美国机械工程师学会(American Society of Mechanical Engineers)”;1884 年成立了“美国电气工程师学会(American Institute of Electrical Engineers)”;1912 年成立了“美国无线电工程师学会(American Institute of Radio Engineers)”,后二者合并为著名的“电气与电子工程师学会(Institute of Electrical and Electronics Engineers, IEEE)”。这些学术团体出版多种专业出版物,召开各种学术会议,因而,它们的建立大大加强与推进了科技情报的交流和传递。由于科学家人数的增加,科技文献的大量涌现,同时,随着印刷技术的不断提高,多种文献载体形式的出现,以及现代化手段在文献处理上的应用,各种文献检索工具也应运而生,并日臻完善,为科技文献的利用和科技情报的传递提供了极为便利的条件。因此,以文献利用为形式的正式传递渠道已成为当代社会情报交流的主要途径。

情报交流的两种方式对科研人员都是非常重要的,两者互为补充。在科研过程中要正确利

用两种渠道获取情报。要具有强烈的情报意识,随时捕捉对自己有用的情报信息,以激发创新的思想火花。

1.2.2 科技文献与科学研究

科学研究是在继承前人或别人成就的基础上进行新知识创造的探索性工作,因而,任何科学研究都必须从收集、掌握和分析科技情报开始。掌握前人或别人已经取得的研究成果,掌握国内外科学研究的现状及其发展,掌握相关学科的研究成果等有关文献资料,其目的就在于在前人或别人研究成果的基础上提出新问题,做出新判断,取得新发现,获得新结论,得到新发展。这就必须在已记载科学技术成果的文献资料中,去获取、消化和掌握自己所需要的优秀研究成果,以开阔视野,扩展思路,接受启迪,并以此为起点去攀登新的科学技术高峰。据调查,日本从事理论科学研究的科技人员收集情报所占用的时间占全部工作的30%,阅读和分析情报的时间为20%,而从事应用技术研究的科技人员用于收集情报的时间为20%,阅读和分析情报的时间为10%。根据我国实际调查统计,从事研究工作的科技人员用在收集、阅读情报上的时间远远低于这个比例。其原因是我国科技人员还没有广泛形成查找利用文献的习惯,或者没有充分认识到科技文献工作对科学研究的重要性。人类发展文明史证明,科学研究工作的开展和科学研究成果的取得,从来就没有离开过文献情报工作。

1.3 文献情报源

科技文献记录和反映着科学技术的成就和水平,包含着无数有用的事实、数据、理论、方法、构思和假设,记载着大量成功与失败的经验教训,反映着一定时代、一定社会条件下科学技术的发展和水平,预示着未来发展的趋势和方向。因此,科技文献是科技人员获取情报的主要来源。

1.3.1 文献的载体形式

文献载体的物质形态和记录知识信息技术的演进与发展,是和社会的认识水平及生产力的状况相适应的。雕版印刷术与造纸技术的发明使文献记载产生了一次飞跃,而随着科学技术的发展,特别是光学技术、电子通信技术及计算机技术的飞速发展,使知识载体发生了新的飞跃。知识载体面临着更加深刻、更加广泛的变革。这是科学技术发展的必然结果。现代文献按载体形式划分有以下几种:

(1) 印刷型文献。包括铅印、油印、胶印、石印等传统的用纸张印刷成的文献。其优点是便于阅读,便于传递,不受时间、地点的限制,可以随时随地直接读取。缺点是存贮密度低,收藏管理要占用较大的空间和人力,很多图书馆已经越来越难以应付不断增长的人藏量,因而部分印刷型文献正逐渐被其他载体所取代。

(2) 缩微型文献。这是一种以感光材料为载体,利用缩微摄影技术为记录手段将印刷文献复制到感光材料上的文献类型,包括缩微胶片(平片)和缩微胶卷。目前最先进的超级缩微胶片或缩微胶卷缩小倍率高达1:22500,一张全息缩微平片可存贮20多万页印刷型文献。

缩微型文献的优点是体积小,存贮密度高,平均可节省95%的存贮面积;便于贮存和保管;价格低,其成本一般只是印刷型的1/10,甚至更低;另外,它的保存时间长,不易损坏和变

质,据报道,国外对缩微品进行的加速试验表明,在恒温恒湿条件下,胶卷和胶片可保存 500 年。

缩微型文献的缺点是必须借助于缩微阅读机,使用不方便,不能同时利用几种文献,而且人的眼睛对屏幕显示容易疲劳。

(3) 机读型文献。机读型文献是通过计算机存贮和阅读的文献类型。它是通过编码和程序设计将文献变为数字语言和机器语言,并输入计算机,存贮在磁盘、磁带、光盘等计算机存贮介质上,建立起相应的文献数据库。阅读是按照一定的指令,再由计算机将存入的信息转换成能直接阅读的自然语言,显示在终端显示器上或由打印机输出。

机读型文献的特点是可以大容量、高密度地贮存信息,并能按指令高速输出文献单元与知识单元。计算机系统与现代通信技术相结合又可以快速的远距离传输文献信息,使得情报传递与情报交流更加便捷,从而超越了地理与时间上的障碍,使世界上出现了即刻就能分享的信息。特别是 INTERNET 和多媒体技术的建立与完善,使机读型文献不仅包括了文字,而且包括了图像、声音等多种媒体形式,并使信息的全球共享成为可能。机读型文献必须借助于计算机才能使用,售价和使用费用较高。

(4) 声像型文献。声像型文献又称直感或视听文献,包括幻灯片、电影片、唱片、录音带、录像带等。它的优点是脱离了传统的单一文字形式,直接记录声音与图像,视听兼备、形象生动,感染力强,可以提高人们理解、吸收与记忆信息的能力。可以能动地再现自然界的种种变化发展,帮助人们认识某些复杂罕见的自然现象。在观察肉眼不易看到的物质运动过程、生态发展,探索物质结构,重复实验过程、形态动作以及语音模拟与再现等方面,具有独特的作用。同时由于它的声音与图像效果,使人闻其声,观其形,可以产生生动逼真的直接感受,是静态的书面文献所不可比拟的。

声像型文献同样需要相应的设备,不如印刷型文献使用方便。

(5) 其他新型载体的文献。随着科学技术的飞速发展,文献载体出现了许多新的类型,例如:直感图书、光盘、立体形象资料等,其中发展较快、应用较广的是光盘和立体形象资料。

光盘存贮技术的发展十分迅速,目前使用较为普遍的光盘存贮量为 650M,现在已出现了容量为 8G 的光盘。光盘存贮技术和计算机技术结合,使检索效率大大提高,可以在几秒钟内查到所需的情报信息。光盘有可能占据未来信息存贮的主导地位。

1.3.2 现代文献的出版类型

(1) 科技图书。图书是历史最悠久的文献类型,一般可分为阅读类和工具类两种。阅读类科技图书包括科技专著、教科书、科普读物等;工具类图书包括词典、百科全书、手册、年鉴等。

科技图书一般是经过作者对大量的资料进行细微研究,选择鉴别,融会贯通,分析、归纳,重新组织编写而成的。所以其内容比较系统、全面、成熟,是可靠的情报来源。科学专著往往还包含作者本人的新资料、新论点,对获得某一专题的全面情报很有参考价值。教科书或综论书则起着综合、积累、传授科技知识,教育和培养科技人才的作用。

工具类图书的查考功能其他类型文献无法比拟的。是快速获取某一数据或事实情报的有效工具。但从时间上看,由于图书出版周期长,内容一般只反映 3~5 年前的研究水平,这就大大削弱了科技图书作为情报源和情报传递的作用。

(2) 科技期刊。科技期刊是指定期或不定期出版的有固定名称的连续出版物。它们有连续

的卷、期或年、月、顺序号,每期载有多个作者的文章。科技期刊的学科广泛、内容新颖、流通面广、报道迅速,能反映当前的科技水平,是传递科技情报、交流学术思想的最基本的文献类型。是目前报道科技情报的最灵活、最迅速的公开形式。

科技期刊论文一直是科技人员最重视的情报来源。美国对数以千计的科学家进行的调查表明,情报需求的68%属于期刊论文;英国电气工程师所用情报源中70%是专业期刊。前苏联科学院对所属单位中化学家利用各类情报源的情况所作的调查表明,科学期刊的利用率为84%。可见,科技期刊是获取科技情报的最主要的来源,据估计,约占全部情报来源的65%~75%,某些学科和专业可能还要大些。

科技期刊还起着汇集其他类型文献的作用,即其他各类型文献常以改写的方式发表在科技期刊上。

(3) 科技报告。是关于某项研究的立项、进展、成果的报告。科技报告时差短,时效性强,其内容高度专业化,大多属于军事技术、国防技术或尖端技术。其中包括各种研究方案的选择和比较,常附有大量数据、图表,涉及的技术手段具体完整,因此具有较高的参考价值。科技报告每篇自成一册,有机构名称和统一编号。但其保密性强,有时不易获得。

(4) 会议文献。科技会议是科技交流的重要渠道。以美国为例,目前每年用于召开科技会议的费用约5000万美元,约占美国政府全部科技情报拨款的1/8。会议文献指在国内外各种学术会议上发表的论文或报告。许多重大的发明或发现往往首先在学术会议上发表。参加学术会议的人员一般具有一定的代表性,大多是本学科的学术带头人或后起之秀,讨论的问题通常都是有关学科当前的重大问题。因此,学术会议基本上可以反映某一学科的最新研究成果及国内外的水平和发展趋势。学术会议兼有直接交流和间接交流两种方式的长处,传递情报比较及时,情报的针对性强。会议文献是重要的情报源。

(5) 专利文献。所谓专利是指用法律保护科技发明创造所有权的制度。专利文献,亦简称专利,指的是专利说明书,是专利申请人向专利机构提交的说明该项发明的目的、技术梗概和专利权限的书面文件。

专利文献包含了丰富的技术情报。专利的范围几乎包括所有的技术领域。具有一定的经济价值,因为特殊用途的新产品或新工艺的发明创造才能获得专利权,所以它是科技人员非常实用而又富有启发性的参考资料和重要的科技情报源。借助于专利文献的长期积累,能了解一个产业部门和一种产品的进化过程。此外,专利文献也是了解国际技术动向的重要依据。

(6) 学位论文。学位论文是指高等学校的本科生、研究生和在研究机构攻读学位的人员为获取学士、硕士或博士学位而撰写的论文。硕士学位论文和博士学位论文是学生在导师指导下,经过一年以上的长期研究完成的、经专家评审和答辩的、带有一定独创性见解的原始研究成果。学位论文探讨的问题比较专一,对问题的来龙去脉阐述比较详细、系统,对科研、生产工作有一定的参考价值。

学位论文一般不公开出版,全部保存在授予学位的单位,只有部分在期刊上摘要发表,所以较难收集。高等学校图书馆及科研单位图书馆应注意收集本单位的学位论文。

(7) 标准文献。标准文献主要指技术标准,是对产品及零部件的质量、规格、性能、成分、生产过程及检验方法所做的技术规定,是人们在设计、生产和检验过程中共同遵守的技术依据。它是一种规章性文件,具有一定的法律约束力。每一件标准都是独立、完整的资料,统一编号。

标准文献的种类很多。按其作用与性质可分为技术标准和管理标准;按其使用范围可分为

国际标准、区域性标准、国家标准、专业标准和企业标准。一个国家的技术标准可以反映一个国家的技术经济政策、生产加工工艺和标准化水平及自然条件、资源等情况。这对全面了解和研究某国的工业发展情况提供重要的情报依据。

随着生产与科学技术的发展,标准的内容不断地修订更新,新标准将逐步取代旧标准。图书馆应根据需要收集标准文献,包括新制订的标准和补充修订的标准文献。

(8) 产品资料。产品资料是生产厂商或推销部门为推销产品而印刷的宣传品和技术资料,包括产品样本、产品目录、产品说明书等。大多数为免费赠送,以配合产品流通进行宣传,扩大产品销售。

产品资料所介绍的是出厂产品的品种、规格、性能、特点、结构、用途及使用维修方法等。技术成熟、数据可靠,常附有外观照片、结构图及规格说明。对造型、设计、改造及技术与设备引进都有较大的参考价值。

1.3.3 文献的系统结构

科技文献按其情报含量与加工程度可以分为一次文献、二次文献和三次文献。

(1) 一次文献。凡是以作者本人的生产与科研工作成果为依据而创作的原始文献,均属一次文献的范畴。它直接记录从事科学研究与生产实践的劳动成果,一般都有详尽具体的学术内容与研究数据,参考和使用价值较大。包括科学专著、期刊论文、科技报告、会议论文、专利文献、学位论文等。

(2) 二次文献。是指经过文献工作者对一次文献加工处理后产生的文献,是对原始文献进行简化与浓缩而形成的工具型文献,如书目、索引、文摘等。二次文献的作用不仅在于报道文献内容,更重要的是可以提供检索原始文献的线索。

(3) 三次文献。三次文献是指在合理利用二次文献的情况下,对一次文献的内容进行分析、归纳、综合、重组而编纂出的文献,如词典、手册、百科全书、年鉴、教科书以及其他综述性和评论性文章等。

科技文献也像任何其他事物一样,具有自身的产生、发展与完善的演变过程。了解文献的产生与演变过程,对于正确地选择与利用文献具有重要的意义。图 1.1 所示的科技文献的链式结构揭示了科技文献的产生与发展的历史性变动。

在这历史性的变动结构中,可以看到每种新知识都有它自身被创立、传递,并被综合到新的知识体系中的动态过程。

在上述的科技文献链中,以“研究发展活动”为起点,按顺时针方向移动,是科技文献由博而约、由分散到集中、由无组织到序列化的过程。随着这个过程的推移,科技文献的有序化程度、可检索性与可获得性递增,但所含情报内容的新颖性与及时性递减。作为链的起点的研究与发展活动,是情报的发源地,又是情报流的目的地。任何科学研究都是从利用情报开始,任何科研人员既是情报的利用者,又是情报的生产者。这种循环往复,推动着人类对自然认识能力的螺旋式上升,推动科学技术的进步,推动人类社会的进步。

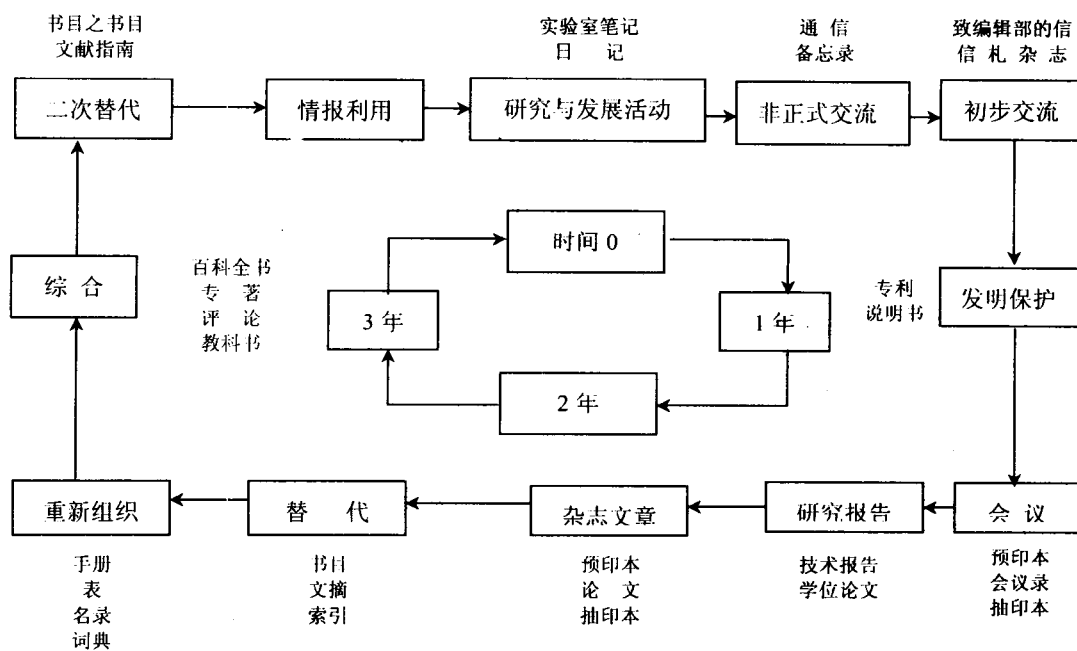


图 1.1 科技文献链

第二章 科技文献检索的基本原理

2.1 文献检索

文献检索是指迅速、准确、无重大遗漏地从浩如烟海的文献海洋中查找出自己所需情报的过程。文献检索亦称为情报检索(Information Retrieval)。

由于情报的本质是知识,科技文献就是科技情报的载体,而科技情报则是科技文献的内容。因此,文献检索是指查找含有情报内容的文献;情报检索则是指查找文献中所含的情报。但实际上,广义的情报检索包含三部分内容,即数据检索、事实检索和文献检索。

数据检索(Data Retrieval),是以数据为对象的检索。例如:要查某一数学公式,某一电子器件的参数,某一材料的成分、性能等。

事实检索(Fact Retrieval),是以特定的事实为检索对象。例如:要查找某一事件发生的时间、地点与过程,某一企业的生产与产品的研制情况,某一高等学校或研究机构的简况等均属事实检索。

文献检索(Document Retrieval),是以文献为检索对象。凡属查找某一主题、某一时代、某一单位、某一著作的有关文献及其出处的,均属文献检索的范畴,例如:要查找有关“碳纤维复合材料的机械性能”有些什么文献,这些文献发表在什么地方?

以上三种检索类型中,文献检索是主体。但三种检索一般都是通过文献进行的,所以,通常称为文献检索。

2.2 文献检索原理

情报检索是情报传递过程的一个重要环节。美国的穆尔斯(Calvin N. Moors)在1950年发表的题为《把情报检索看作是时间性的通讯》的论文中指出:“情报检索是一种时间性的通讯形式,它与点对点的通讯相区别。”“此种通讯是时间性的,在时间上从一个时刻通往较晚的时刻,而在空间上可能还在同一地点。”这就是说情报接收与提供之间有一段时间的延迟,首先要将接收来的情报存贮起来,以供日后检索。所以,情报检索实际包含存贮与检索两个过程。

存贮过程——对具有情报价值的大量的原始文献进行分析,提取其具有检索意义的外表特征与内部特征,著录下来,给以一定的标识符号,按一定的规则排列起来,组成手工检索工具或建立计算机文档,这就是存贮过程。

检索过程——对所需情报内容进行分析、标引处理,选取提问特征,将其转换成检索标识,再通过检索工具或计算机文档,按一定的方法找出所需的情报或文献线索。

综上所述,文献检索的基本原理如图2.1所示:检索者的检索提问标识与存贮在检索工具或文档中的文献特征标识进行匹配对比,如果能取得一致,或文献特征标识包含检索标识时,即为检索命中,具有该文献特征标识的文献著录就从检索工具或文档中输出。检索者可据

此进行判断、筛选,进而索取原始文献。如果是手工检索,这个过程基本是手翻、眼看、大脑进行思考、推理和判断的过程,若为计算机检索,匹配对比则是在检索标识输入后,有计算机自动进行。

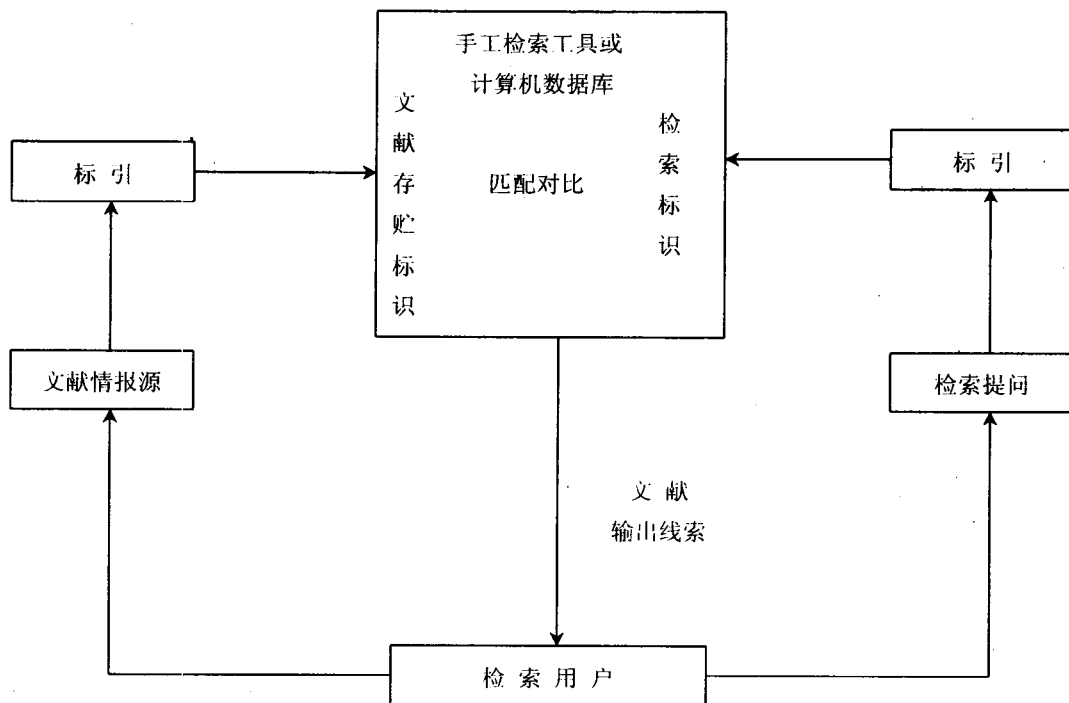


图 2.1 文献检索系统示意图

2.3 检索语言

情报的存贮和检索的主体都是人。存贮标引者与检索者的专业知识、工作经历、语言习惯有所不同,如果没有共同的语言就不能保证存贮与检索的一致性,必然给检索造成极大的困难。因而,就需要创制一定的语言,在情报存贮时用它来描述文献的内容特征与外表特征,在情报检索时也用它表达。这种把文献的存贮与检索联系起来,使人们取得共同理解,实现交流的人工语言称为检索语言。

检索语言是根据情报检索的需要而编制的一种人工语言,是情报检索系统的重要组成部分,在情报检索系统中起着语言保障的作用,是沟通情报存贮和检索两个过程,标引人员和检索人员双方思想的桥梁。如果没有检索语言作为标引人员和检索人员的共同语言,就很难使标引人员对情报内容的表达(标引用语)和检索人员对相同内容的情报需要的表达(检索用语)取得一致,情报检索也就不可能顺利实现,甚至根本不能实现。由于其使用的场合不同,常使用不同的名称,例如,在存贮的过程中用来标引文献,就称为标引语言;用以编制索引就称为索引语言;用来检索文献则称为检索语言。

2.3.1 检索语言的类型

检索语言按描述文献的有关特征,可分为描述文献外表特征的语言和描述文献的内容特征的语言,如图 2.2 所示。

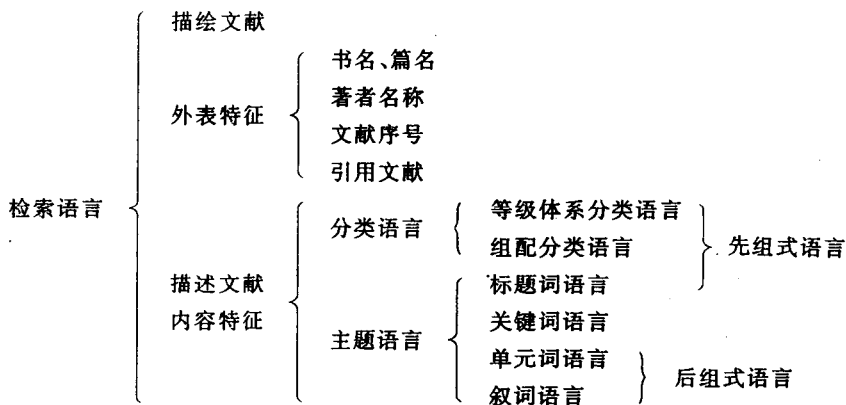


图 2.2 检索语言的类型

描述文献外表特征的语言是以文献上标明的、显而易见的特征,如篇名、著者姓名、文献序号等作为文献标识与检索依据,简单明了,因此,人们对检索语言的研究主要是描述文件内容特征的语言。

2.3.2 分类语言

“类”是指具有许多共同属性事物的集合。每一个事物都有多种属性,用其某一种属性作为划分依据来对一类事物进行划分就叫做分类。分类是人类逻辑思维的一种最基本的形式。人们总是习惯于对世界上的事物进行分类,而这种分类的方法在文献处理上得到了广泛的应用。

分类语言是按知识的学科属性,运用逻辑分类的原理将学科概念作多层次划分,从而形成一个能充分体现事物的隶属、平行、派生关系的严格有序的直线性知识门类的等级制体系。具体地说,它是以学科分类为基础,按照概念划分的原理,将知识概念从具体到一般、从简单到复杂、从低级到高级逐级划分,每划分一次就形成一批并列的概念——下位概念,它们同属于一个被划分的概念——上位概念。例如,“固体力学”这个类可以划分为“材料力学”、“结构力学”、“弹性力学”、“塑性力学”等子类。“固体力学”属被划分的类,即上位类(上位概念),它所划分出的几个子类,即下位类(下位概念)。这几个下位概念体现的是平行关系,而上下位类之间则是隶属关系。每个类目都用分类号作为标识,每个分类号就代表特定的知识概念。体现这种分类体系的就是图书分类法(表)。

分类法种类很多,比较影响的有《中国图书馆图书分类法》、《国际十进制分类法》(Universal Decimal Classification, UDC)、《杜威十进制分类法》(Dewey Decimal Classification, DDC)、《美国国会图书馆分类法》(Library of Congress Classification, LCC)、《冒号分类法》(Colon Classification, CC)等。《中国图书馆图书分类法》(简称《中图法》)作为我国文献分类标引工作的国家标准,被我国图书情报单位广泛应用。其大类体系结构及类目布局如下:

A 马克思主义、列宁主义、毛泽东思想、邓小平理论	TD 矿业技术
B 哲学、宗教	TE 石油、天然气工业
C 社会科学总论	TF 冶金工业
D 政治、法律	TG 金属学与金属工艺
E 军事	TH 机械、仪表工业
F 经济	TJ 武器工业
G 文化、科学、教育、体育	TK 能源与动力工程
H 语言、文字	TL 原子能技术
I 文学	TM 电工技术
J 艺术	TN 无线电电子学、电信技术
K 历史、地理	TP 自动化技术、计算机技术
N 自然科学总论	TQ 化学工业
O 数理科学和化学	TS 轻工业、手工业
P 天文学、地球科学	TU 建筑科学
Q 生物科学	TV 水利工程
R 医药、卫生	U 交通运输
S 农业科学	V 航空、航天
T 工业技术	X 环境科学、安全科学
TB 一般工业技术	Z 综合性图书

2.3.3 主题语言

主题语言是一种描述性语言,它用语词直接表达情报的主题,这些语词就是表达主题概念的标识。将这些作标识的语词按字顺排列并使用参照系统来间接表达各种概念之间的关系,这就是主题语言。

主题语言所使用的语词统称为主题词。主题词是用以表达描述文献主题概念的名词术语。这些名词术语,来自自然语言,多半经过规范化处理,形成主题词表,作为标引与检索的依据。

根据语词的选词原则、组配方式、规范方法等,主题词语言可分为标题词语言、关键词语言、单元词语言和叙词语言。

标题词语言是以标题词作为文献内容标识和检索依据的一种主题法语言。标题词是来自自然语言中比较定型事物的名称,并经规范化处理的、能表达文献主题内容的词、词组或短语。在编制标题词表时,事先将一级标题词和二级标题词加以固定组配,所以它属于定组式先组配检索语言。

关键词语言是以关键词作为文献内容标识和检索依据的一种主题法语言。关键词是选自文献题目、文摘乃至正文中的具有实质意义的语词,它没有固定的词表,对于同一事物概念,不同著者,甚至同一著者在其不同的著作中用词也会有所不同。因此,关键词是一种未经规范化的自然语言词汇。

单元词语言是以单元词作为文献内容标识和检索依据的一种主题法语言。单元词是从文献中抽选出来的,从字面上不可再分割的最基本的概念单元的词汇。检索时,可根据情报提问

内容选用适当的单元词进行组配,故它是一种后组式检索语言。

叙词语言是目前主题语言的高级形式。叙词是从文献内容中抽选出来的、从概念上不可再分的基本概念单元词。检索时,利用这些概念单元进行组配,以表达一个复杂的概念。叙词语言是近代用途较广,既适用于手工检索,又适应于计算机检索的后组式检索语言。

情报检索语言的种类虽多,但常用的手工检索工具中所使用的检索语言,也只有体系分类语言、标题词语言、关键词语言和叙词语言。

2.4 检索工具

2.4.1 检索工具的定义

检索工具是指用以存贮、报道和检索文献线索的工具。

如前所述,科技文献的数量庞大、交叉重复、高度分散,而且增长速度越来越快,这就给文献的利用带来了极大的困难。为此,人们在长期的实践过程中,创造出了一种压缩文献的方法,即提取文献的特征,加以标引,并按一定的顺序排列起来,从而形成一种二次文献,便于人们用较少的时间从不同的角度获取大量文献线索,这种二次文献就是检索工具。

检索工具按处理文献的手段可分为手工检索工具和机械检索工具。手工检索工具是指使用手工方式来处理和查找文献的书本式检索类出版物。这是在长期历史实践中形成的传统检索工具,包括目录、题录、索引、文摘等。机械检索工具是利用机械、光电等手段帮助完成检索的工具,通常称为检索系统,目前主要指计算机检索系统,它代表了文献信息检索的发展趋势,有着广阔的前途。

2.4.2 传统检索工具的基本类型

在长期的检索实践过程中,为了适应不同的检索需要,逐渐形成了传统检索工具的几种基本类型,即目录、题录、索引、文摘。

1. 目录(Bibliography, Catalogue)

目录一般以整本的图书、期刊等作为报道单元,注明具体的出版或收藏单位,仅著录出版物的外表特征,按类编排,主要用于查找出版物的出版或收藏单位,根据编制目的和其社会职能,目录可以分为国家目录、馆藏目录、联合目录和出版社与书商目录等。

2. 题录(Title)

题录是在目录的基础上发展起来的、以出版物中的“篇”作为著录单元的检索工具。一篇文献作为一条款目,如期刊中的一篇论文。因此,在揭示文献的内容上比目录更进一步。题录的特点是“快”和“全”。题录一般不做过多的加工,不做内容摘要,仅列出篇名、著者,出处。

由于题录和索引在功能上有相近之处,有时人们把题录也称为索引,例如我国的《全国报刊索引》就是典型的题录型检索工具。

3. 索引(Index)

索引是将散见于文献中的知识单元,如题名、人名、地名、主题词、分子式和各种号码(如专利号、报告号等)分别摘出,并按一定的顺序排列,注明相应的出处,以便人们查找的一种工具。

4. 文摘(Abstract)

文摘是将论文或专著的内容加以浓缩,以最精练最概括的文字报道文献主题、方法和结论,即文章的内容摘要。将描述文献外表特征的著录加上文摘,按一定的顺序排列起来,即形成文摘型检索刊物,常简称为文摘。其实质就是题录加上内容摘要。

(1) 指示性文摘(Indicative Abstracts):文摘内容较简单,一般在 100 字左右,简明扼要地介绍文献的主题范围、研究方法、结论、用途等。不涉及具体技术内容,以使读者对文摘内容不致产生误解为原则,所以也有人将其称为简介。

(2) 报道性文摘(Information Abstracts):是对原文内容的高度浓缩。报道原文的论点、方法、设备、结论、具体数据等。我国国家标准规定,一般字数为 200~300 字(西文为 250 个实词),如有特殊需要字数可略多。这种文摘含量大,参考价值高。

文摘类刊物中,有时是指示性文摘、报道性文摘和题录并用。

文摘类检索刊物主要由文摘和索引两大部分组成。文摘部分主要起报道作用,大多按分类或主题编排,构成检索刊物的正文。索引部分起检索作用。所以,检索的种类是否全面,编制是否合理,使用是否简便,直接影响整个检索工具的检索效率,是衡量一个文摘刊物质量的重要标志。

2.4.3 科技文献检索工具的一般结构

一部完整的文献检索工具一般由五部分组成,如图 2.3 所示。

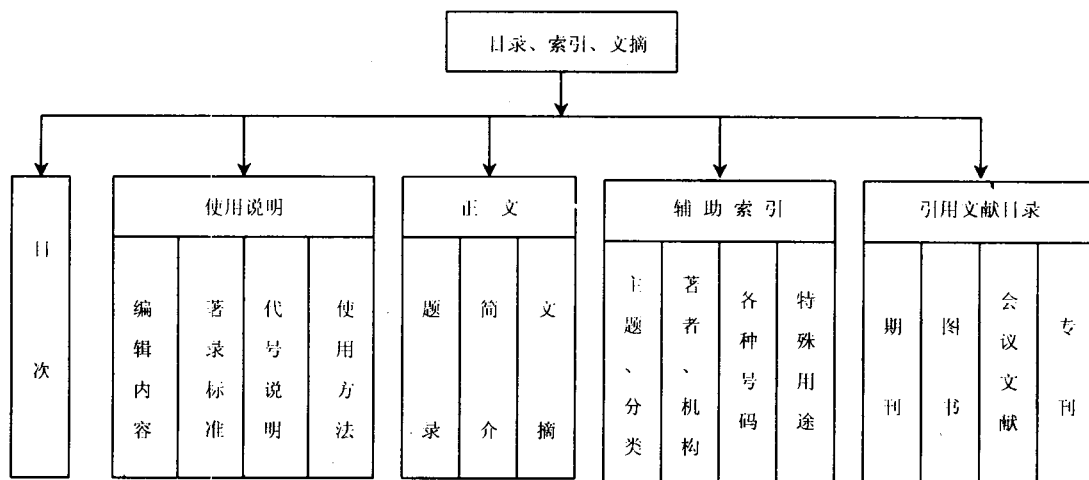


图 2.3 检索工具一般结构

1. 使用说明

为利用者编写的必要的指导,一般包括编制目的、收录范围、著录格式、代号说明及注意事项等,使用前应认真阅读。

2. 目次表

检索工具的正文一般都按分类组织编排,所以前面都有详简不等的目次表,作为从分类途径查找文献的依据。

3. 正文

这是检索工具的主体部分。存贮在检索工具中的内容不是原始文献,而是描述文献的外表